Jurnal Teknologi

Full paper

A Probabilistic Individual-based Model for Infectious Diseases Outbreaks

Pierpaolo Vittorini[a*], Antonella Villani[a], Ferdinando di Orio[a]

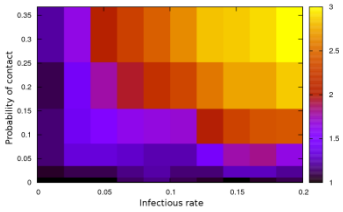
Dep. of Life, Health and Environmental Sciences, University of L'Aquila, 67100 L'Aquila, Italy

*Corresponding author: pierpaolo.vittorini@univaq.it

Article history

Received :11 September 2012
Received in revised form :
21 February 2013
Accepted :15 April 2013

Graphical abstract



Abstract

The mathematical modelling of infectious diseases is a large research area with a wide literature. In the recent past, most of the scientific contributions focused on compartmental models. However, the increasing computing power is pushing towards the development of individual models that consider the disease transmission and evolution at a very fine-grained level. In the paper, the authors give a short state of the art of compartmental models, summarise one of the most know individual models, and describe both a generalization and a simulation algorithm.

Keywords: Computational epidemiology; infectious diseases; compartmental models; high-resolution models, computer simulations

1.0 INTRODUCTION

Computational epidemiology is a multidisciplinary field that brings together diverse contributions coming from computer science, mathematics, statistics, geographic information science and public health, so to help epidemiologists in their studies concerning e.g. the evolution of epidemics.

In such a context, the mathematical modelling of infectious diseases has a long tradition [10, 12]. Currently, different approaches exist: compartmental models based on differential equations [8, 9], ad-hoc models for the contact process [11, 1], or individual-based models [7, 4, 14].

The paper starts describing the compartmental models, then discusses a relevant individual-based model (i.e., the Eubank model [7]), and delves into a recent extension [14] by presenting a further generalisation and diverse stopping criteria.

2.0 COMPARTMENTAL VS INDIVIDUAL MODELS

Compartmental models divide the population into compartments (groups of subjects with homogeneous characteristics) and describe the variation of the number of subject that moves from one compartment to another through differential equations.
For instance, the SIR model uses the following compartments: (i) S: susceptible, (ii) I: infected, and (iii) R: recovered. Furthermore, let:

- β be the contact rate, i.e. the rate of becoming infectious by contacting another susceptible subject;
- γ be the recovery rate, i.e. the rate of recovering from an infection.

The variations in the time of the number of susceptible, infections and recovered individuals (by also assuming that the number of deaths are equivalent to the number of births) are described by the following equations:

$$\frac{dS}{dt} = -\beta IS;\ \frac{dI}{dt} = \beta IS - \gamma I;\ \frac{dR}{dt} = \gamma I \quad (1)$$

According to (1), in the time, the number of susceptible individuals S decreases as of the infections (calculated in terms of the contact rate, number of susceptible and infected individuals), and the number of infected individuals I increases of the previous quantity and decreases as of the individuals that recovers from the infections (calculated in terms of the recovery rate and the number of infected individuals).

The SIR model can be extended by including, i.e. the birth/death rate, vaccinations, and by even adding further compartments, thus leading to more complex models (e.g. the SIER model).

It is worth stressing that the compartmental models are valid only in case of sufficiently large populations. Since a large population comprises of many different individuals in various fields, the diversity is reduced to a few key characteristics which are relevant to the infection under consideration, thus smoothing over the differences of each individual (e.g. specific behaviours, movements).

Instead, the so-called individual-based models try to take into account the specificities of the individuals composing the population and the way in which each individual can differently contract the disease or infect another individual. In particular, such approaches:

- build up a social network that realistically estimates the way in which every individual may contact other individuals;
- divide the epidemic process in terms of two sub-models, called between-host disease transmission and within-host disease progression. The first takes care of describing the disease transmission from one individual to another, the second takes care of describing the disease progression within each individual.

One of the most known individual-based models is the one proposed by Eubank at al [7]. Such a model has the following characteristics:

- the social network is build through a software agent called TRANSIM [13, 3], that is able estimate the movements of each individual in a urban area;
- the between-host disease transmission model is based on bipartite graphs, in which the two classes of vertices are persons and locations, and the edges connects individuals to locations with a label that specifies the period of time in which an individual visited a location. For instance, the graph depicted in Figure 1 shows person p_2 in L_1 from 8:00 to 9:00, in L_2 from 10:00 to 12:00, and in L_3 from 9:00 to 10:00;
- the within-host disease progression is modelled as follows: an individual becomes infected if in the same place of an infected individual for more than a certain period of time, and depending on the disease infectious rate.

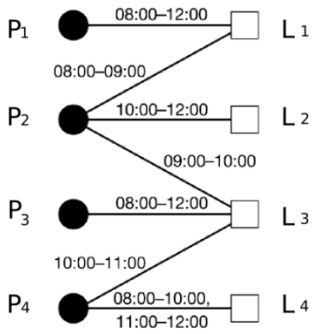


Figure 1 Example of bipartite graph estimated by TRANSIM

It is worth noting that Eubank *et al*. estimated the social network for Portland (Oregon, USA), simulated a smallpox epidemics, and demonstrated – differently from the compartmental models that would have suggested mass vaccination – that the epidemics could have been better controlled through focused quarantine and vaccinations, combined with early detection.

To understand such a result, let us take into account Figure 2. Each node of the graph is an individual, and each edge between nodes represents that the two individuals are in contact each other, and thus may transmit each other an infectious disease. Let us suppose that the epidemic starts in an individual belonging to the area enclosed in the dashed box. For the epidemic to spread, it has to pass through the "bottleneck". Therefore, if we could have (i) an early detection system able to signal the presence of a possible outbreak and (ii) the social network of the population under analysis, we could stop the outbreak by putting into quarantine the sub-graph enclosed in the dotted box, and by targeted vaccination.

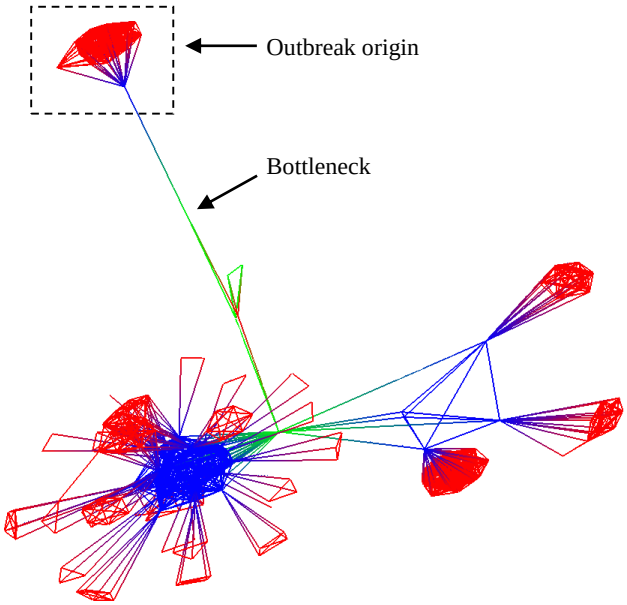


Figure 2 Sample social network

3.0 THE EXTENDED MODEL

The Eubank model uses exact movements of people (in order to properly label the arcs) and does not take into account the evolution of infectious diseases spread by vectors (the only classes of nodes are in fact people and places).
In this section, we summarize an extension of this model [14], that addresses the two limitations above as follows. In the between-host disease transmission model we introduce: (i) probability functions that captures the uncertainty about the movements of people and (ii) a further class of nodes representing vectors.

Furthermore, we make use of probabilistic timed automata [5, 6] to model the within-host progression.

3.1 The Mathematical Basis

3.1.1 Between-host Transmission Model

The between-host model is a tripartite graph with three types of vertices that represent people, locations and vectors. A person is identified with the notation p, a vector with v, a place with l. Let us assume respectively N, V and L be the number of individuals, vectors and places. Furthermore, at each discrete time t, a person p (or a vector v) is in only one place.

The edge that connects a person p (or a vector v) to a place l is labelled by a probability function $f_{p,l}(t)$ (or $f_{v,l}(t)$ for a vector), which represents the probability that, at time t, person p (or vector v) is in location l.

A person can contract the disease either by means of a contact with an infected person p' located in the same place of p, or due to the presence of a vector v. In addition, the individual may contract the disease or not according to a certain probability, which may depend on various factors (e.g. by the immune resources of the subject, the specificities of the disease, the place in which it may be contracted, etc). Therefore, the probability that

subject p becomes infected because of the infected individual p', in l, at time t, is given by:

$$f_{p,l,p'}(t) = \gamma_{p,l,p'}(t) \cdot f_{p,l}(t) \cdot f_{p',l}(t) \qquad (2)$$

where $\gamma_{p,l,p'}(t)$ is the probability of disease transmission from p to p' in location l.
Similarly, if the disease is transmitted by vectors, the probability for person p to contract the disease due to the presence of vector v in location l in time t is given by:

$$f_{p,l,v}(t) = \tau_{p,l,v}(t) \cdot f_{p,l}(t) \cdot f_{v,l}(t) \qquad (3)$$

where $\tau_{p,l,v}(t)$ is the probability for person p to contract the disease from vector v in location l. It is worth remarking that the model can take into account also aggregations of persons and vectors (e.g. a swarm of mosquito). Further details on this can be found in [14].

3.1.2 Within-host Progression Model

The within-host disease evolution is modelled as a finite state automata with probabilistic transitions [5], in a manner similar to that proposed in the work of Dodds & Watts [6]. The states of the automata represent the state of health of subject p (e.g., healthy, infected or dead), while the edges that connect the states are labelled by probability functions $f_{p,s,s'}(t)$ that describe, in the time, the probability for subject p to move from state s state to state s'.

For instance, Figure 3 shows a sample automata for a three state disease progression (healthy, infected, dead). In particular, from the healthy state, an individual may become infected as of equations (2) or (3) above. Then, the infected individual can heal with probability function $h_{p,d}(t)$, or die with probability function $d_{p,d}(t)$, or remain infected with the remaining probability.

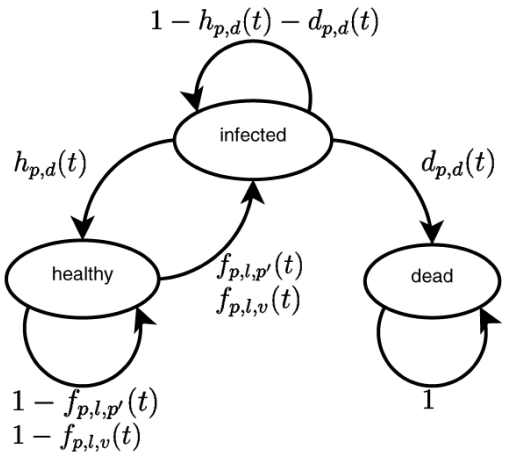


Figure 3 Example of within-host disease progression

3.2 The Simulative Process

An ad-hoc simulation exploits the aforementioned between-host transmission and within-host progression models. The simulation summarised in Algorithm 1 starts at time t_0 and ends at time t_1, with a discrete time interval Δt. It simulates firstly the between-host disease transmission (lines 4–11), then the within-host disease progression (lines 13–21). In summary:

- Concerning the between-host disease transmission, the algorithm cycles over all locations that (at time t) has an infected individual on them. Then, cycles over all healthy persons in the same locations. By using equations (2) and (3), calculates the probability that each healthy person may become infected. Finally, it extracts a random number, and if the calculated probability is larger than the random number, we assume that the transmission took place;
- Concerning the within-host disease progression, the algorithm cycles over all non-healthy people and examines all possible state evolutions of the related probabilistic timed automata. Similarly, by using a random number, the disease evolves accordingly.

Algorithm 1 Iterative simulation

```
1  t = t0
2  while t ≤ t1 do
3    // BETWEEN-HOST DISEASE TRANSMISSION
4    foreach location l with an infected person p' or
               vector v do
5      foreach healthy person p with an edge in l do
6        prob = f_{p,l,p'}(t) or f_{p,l,v}(t)
7        r = random [0,1)
8        if (prob > r) then
9          p becomes infected
10     end
11   end
12   // WITHIN-HOST DISEASE PROGRESSION
13   foreach non-healthy person p do
14     let s be the state of p
15     foreach state s' connected to s do
16       prob = f_{p,s,s'}(t);
17       r = random [0,1)
18       if (prob > r) then
19         p is in state s'
20     end
21   end
22   t = t + Δt
23 end while
```

It is clear that repeated executions of Algorithm 1, even on the same scenario, can produce different results, since both the between-host transmissions and the within-host progressions are influenced by the extraction of a random number. Therefore, similarly to the well-know Montecarlo simulations, we must repeat the algorithm as long as an adequate stop criterion is fulfilled.

It is worth noting the impact of such repetitions. During the iterations, different scenarios are generated. By comparing them, we could isolate the worst case (e.g. when we have the largest number of deaths), the best case, and the average one, so to have further opportunities to decide the best preventive and/or healing action.

Hereafter, two different stop criterions are presented.

1. The first criterion consists in stopping when the average number of persons in a certain state (e.g., infected, dead) is enough accurate wrt a given error;
2. The second criterion, instead, consists in stopping when the number of times in which *each* person is in a certain state (e.g., healthy, infected) is enough accurate wrt a given error.

The first criterion is preferable when the researcher is interested in studying the epidemics at the high level. When the researcher is interested in stabilising also the results of each individual, the second criterion is instead more proper.

3.2.1 Stabilisation on the Number of Individuals in a Certain State

According to the central limit theorem, the algorithm is stopped when the following condition occurs:

$$2 \cdot x_{\alpha/2} \cdot \sqrt{\frac{S_N^2(t)}{N}} < \varepsilon \qquad (4)$$

where N is the current number of iterations, $S^2_N(t)$ is the variance of the number of subjects belonging to the desired state (e.g., infected, dead), 1−α is the confidence level, $x_{\alpha/2}$ is chosen so that $\int_{-\infty}^{x_{\alpha/2}} g(t)dt = 1 - \alpha/2$, g(t) is the normal distribution, and ε is the acceptable error level.

3.2.2 Stabilisation on the Number of Times in which Each Individual is in a Certain State

Let us focus on the healthy state. Similar considerations applies to any other state. For any person p, we define

$$T_p(N) = \text{number of time in which the person p is healthy in the simulation N,}$$

and T_p^* is the associated empirical mean, and T_p the true value. In order to found the simultaneity interval confidence for T_p, thanks to the Central Limit Theorem, similarly as in §3.2.1, we consider that the algorithm must be stopped when the following condition occurs *for all* persons p:

$$2 \cdot x_{\alpha/2} \cdot \sqrt{\frac{T^2_{p,N}(N)}{N}} < \varepsilon \qquad (5)$$

where N is the current number of iterations, $T^2_{p,N}(N)$ is the variance of T_p^*, 1−α is the confidence level, $x_{\alpha/2}$ is chosen so that $\int_{-\infty}^{x_{\alpha/2}} g(t)dt = 1 - \alpha/2$, g(t) is the normal distribution, and ε is the acceptable error level.

4.0 CASE STUDY

The case study describes the effect of probability in the transmission of the disease, in comparison with the model proposed by Eubank *et al*. In particular, let us refer to the following scenario. Let us suppose that an infected person visits a gym from 12:00 to 13:00, and 50 healthy individuals arrive at 13:00 and go off at 14:00, each for a different location. Furthermore, let us assume that the disease is such as to have an infectious rate of 10%, without any incubation period.

In the model of Eubank et. al., the disease transmission occurs only if one can assume a contact for a more than a certain period of time. Therefore, in the aforementioned scenario, given the absence of any contact between the infected and the healthy individuals, the approach of Eubank at al. deduces the impossibility of the outbreak.

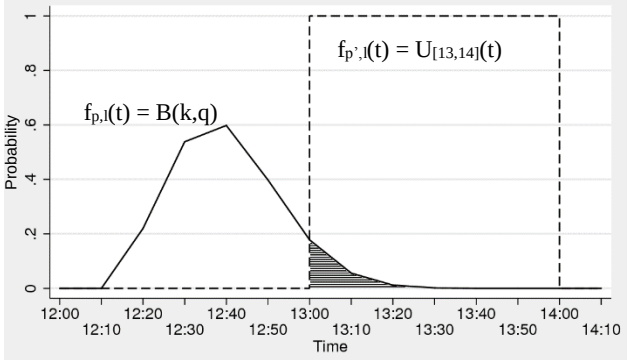


Figure 4 Binomial B(k,q) and uniform U[13,14](t) variables

However, let us suppose that we could introduce uncertainty about the time in which the infected person visits the gym, and in particular let us suppose that we could model this uncertainty with a Binomial random variable with parameters k = 11 and k · q = 2, as shown in Figure 4.

$f_{p,l}(t) = B(k,q)$
$f_{p',l}(t) = U[13,14](t)$
$\gamma_{p,l,p'}(t) = 0.1$
$\tau_{p,l,v}(t) = 0$

Accordingly, equations (2) and (3) become:

$f_{p,l,p'}(t) = 0.1 \cdot B(k,q) \cdot U_{[13,14]}(t)$
$f_{p,l,v}(t) = 0$

Furthermore, the equations for the within-host disease progression (as of the example model described in §3.1.2) are as follows:

$$h_{p,d(t)} = d_{p,d}(t) = 0$$

i.e., an individual, when becomes infected, cannot die nor heal.

In such a context, it is now possible for the infected individual to contact the healthy persons and thus to give birth to an epidemic. In particular, the area behind the tail of the binomial distribution from 13:00 to 13:40 represents such a probability.

Figure 5 shows the results of the simulation applied on the case study described above, by using the first stop criterion. In particular, the graph shows the best (dotted line), worst (point-dashed line) and average cases (dashed line). As can be noticed, the best case corresponds to the results of the Eubank *et al*.'s model, i.e., the epidemic does not spread. The worst case is that the number of infections quickly increases until all individuals become infected. Finally, the average case consists in a slower increase, until nine infections are detected, with a confidence interval of plus/minus one infection.

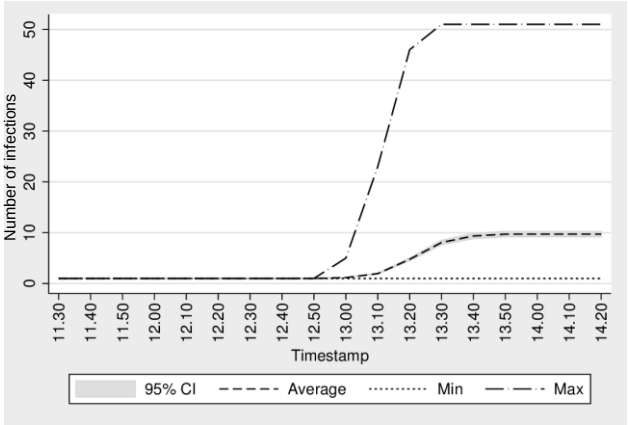


Figure 5 Case study results

5.0 CONCLUSIONS

The paper summarises few approaches that take into account disease outbreaks, starting with the traditional mathematical approaches (e.g. SIR) and ending with individual-based ones. The advantages of individual-based approaches were essentially connected with the ability of gaining better insights into the epidemics. Nevertheless, there are several issues that currently limit their application in real scenarios:

- The creation of the social network is an extremely complex process. Although e.g. a software as TRANSIMS can estimate the contact network, it is limited to only urban areas. However, the possibility of approximating portions of the population by adequately modifying the probability distributions is a possible approach [14];
- The execution of the simulation is an extremely expensive process from a computational point of view, and the resources required are very large. In this direction there are proposals that attempt either to reduce the algorithmic complexity of the simulation [14] or to use parallel architectures [2].

References

[1] Alloran, M., I. Longini Jr, A. Nizam, and Y. Yang. 2002. Containing bioterrorist smallpox. *Science.* 298: 1428–1432.

[2] Barrett, C., K. Bisset, S. Eubank, X. Feng, and M. Marathe. 2008. 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. IEEE Press, Piscataway, NJ, USA. 1–12.

[3] Beckman, R., K. Berkbigler, K. Bisset, B. Bush, S. Eubank, K. Henson, J. Hurford, D. Kubicek, M. Marathe, P. Romero, J. Smith, L. Smith, P. Speckman, P. Stretz, G. Thayer, E. Eeckhout, C. Barrett, and M. Williams. 2003. TRANSIMS, chap. Chapter 3. Los Alamos National Laboratory.

[4] Carley, K., D. Fridsma, E. Casman, A. Yahja, N. Altman, L. Chen, B. Kaminsky, and D. Nave. 2006. BioWar: scalable agent-based model of bioattacks. Systems, Man and Cybernetics, Part A. *IEEE Transactions on Systems and Humans*. 36(2): 252–265.

[5] Beaquier, D. 2003. On Probabilistic Timed Automata. *Theoretical Computer Science*. 292(1): 65–84.

[6] Dodds, P., and J. Watts. 2005. A Generalized Model of Social and Biological Contagion. *Journal of Theoretical Biology*. 232: 587–604.

[7] Eubank, S., H. Guclu, A. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkal, and N. Wang. 2004. Modelling Disease Outbreaks in Realistic Urban Social Networks. *Nature.* 429(6988): 180–4.

[8] Godfrey, K. 1983. *Compartmental Models and Their Application*. Academic Press.

[9] Grassly, N., and C. Fraser. 2008. Mathematical Models of Infectious Disease Transmission. *Nat. Rev. Microbiol*. 6(6): 477–487.

[10] Hethcote, H. 2000. The Mathematics of Infectious Disease. *SIAM Review*. 42: 599–653.

[11] Keeling, M. 1999. The Effects of Local Spatial Structure on Epidemiological Invasions. Proceedings of the Royal Society of London B. *Biological Sciences.* 266: 859–867.

[12] Kenrad, N., and C. Masters. 2006. *Infectious Disease Epidemiology: Theory and Practice.* Jones & Bartlett Publishers.

[13] Rickert, M., and K. Nagel. 2001. Dynamic traffic assignment on parallel computers in TRANSIMS. *Future Gener. Comput. Syst*. 17(5): 637–648.

[14] Vittorini, P., A. Villani, and F. Di Orio. 2010. An Individual-based Networked Model with Probabilistic Relocation of People and Vectors Among Locations for Simulating the Spread of Infectious Diseases. *Journal of Biological Systems.* 18(04): 847–866.