EPIDEMIC INDIVIDUAL-BASED MODELS APPLIED IN RANDOM AND SCALE-FREE NETWORKS

Christofer Roque Ribeiro SILVA[1]
Alexandre Celestino Leite ALMEIDA[2]
Rodrigo Tomás Nogueira CARDOSO[3]
Ricardo Hiroshi Caldeira TAKAHASHI[4]

■ ABSTRACT: This work proposes a version of the Individual-Based Model (IBM) that converges, on average, to the result of the SIR (Susceptible-Infected-Recovered) model, and studies the effect of this IBM in two types of networks: random and scale-free. A numerical computational case study is considered, using large scale networks implemented by an efficient framework. Statistical tests are performed to show the similarities and differences between the network models and the deterministic model taken as a baseline. Simulation results verify that different network topologies alter the behavior of the epidemic propagation in the following aspects: temporal evolution, basal reproducibility and the number of infected in the final.
Keywords: Individual-based models; random networks; scale-free networks.

1 Introduction

Mathematical Epidemiology is an area of science that seeks to more effectively model the behavior of epidemics using mathematical and computational tools. In this sense, several models have been proposed with the purpose of describing and quantifying, in the best possible way, the process of propagation of epidemics. Generally, these models can be framed in two distinct classes: continuous-time

[1]Centro Federal de Educação Tecnológica de Minas Gerais - CEFET-MG, Programa de Modelagem Matemática e Computacional, 30510-000, Belo Horizonte, Brasil. E-mail: *christofer.roque@gmail.com*
[2]Universidade Federal de São João Del Rey - UFSJ, Campus Alto Paraopeba, 36420-000, Departamento de Física e Matemática, Ouro Branco, Brasil. E-mail: *celestino@ufsj.edu.br*
[3]Centro Federal de Educação Tecnológica de Minas Gerais - CEFET-MG, Departamento de Matemática, 30510-000, Belo Horizonte, Brasil. E-mail: *rodrigoc@des.cefetmg.br*
[4]Universidade Federal de Minas Gerais - UFMG, Departamento de Matemática, 31270-90, Belo Horizonte, Brasil. E-mail: *taka@mat.ufmg.br*

models, based on ordinary or partial differential equations, being the SIR model (KERMACK; MCKENDRICK, 1927) the main representative one; and discrete-time models, with the purpose of simulating specific characteristics of the epidemic behavior, being the Individual-Based Model - IBM (KEELING; GRENFELL, 2000; NEPOMUCENO; RESENDE; LACERDA, 2018) an example.

The study of networks (or graphs) is a multidisciplinary area that aims to understand how the objects connect and what are the implications of such connections. Several previous works analyzed, analytically and using simulations, classic epidemiological models and their extensions for networks (NOWZARI; PRECIADO; PAPPAS, 2015; NADA, a; PASTOR-SATORRAS; VESPIGNANI, 2002) or under the point-of-view of Agent-Based Models (MACAL; NORTH, 2010; ABAR et al., 2017). The purpose of including networks is to represent a process of propagation of epidemics that most closely represents what occurs in different realities. For example, the propagation of epidemics may be different in human communities of large cities, where some individuals daily just move around their houses and others have to travel long distances, and in small cities, in which few individuals relate to a few contacts.

Some works have compared the classical deterministic SIR and its stochastic versions (MAGAL; RUAN, 2014) and other works focus on the construction of agent-based models for studying the effects of the spread of some disease in large scale, such as influenza epidemic, respiratory infections, and bovine brucellosis (RAKOWSKI et al., 2010; GE et al., 2018; NEPOMUCENO et al., 2018). A theoretical investigation on how topology affects the spread of an epidemic in several important network topologies has also been performed (GANESH; MASSOULIÉ; TOWSLEY, 2005; HE; MIEGHEM, 2018). These articles state that the computational simulation of complex networks leads to a much greater computational intensity and difficulty in analyzing the significance of parameters. In fact, all network-based simulations are limited, since there is no simple way to ascertain the sensitivity of the epidemiological results to the details of the network structure (KISS et al., 2017). Generally, small networks can be analyzed graphically, allowing visual understanding of their nature, but increasing the size of the network the graphical analysis becomes very difficult, and so statistics become relevant (NEWMAN, 2003).

The behavior of epidemics in societies with different structures has been modeled by networks in several works (PASTOR-SATORRAS; VESPIGNANI, 2001a, 2001b; BARTHELEMY et al., 2004; HWANG et al., 2005; SERRANO; BOGUNA, 2006; EAMES, 2008; MILLER, 2009; NOEL et al., 2009; NEWMAN, 2009; BADHAM; STOCKER, 2010). For example, the article (YOUSSEF; SCOGLIO, 2011) describes an approach based on individuals to model the spread of diseases in networks. In this approach, each individual has a probability of being in susceptible, infected or recovered states. This model is built using a continuous-time Markov chain, able to evaluate probabilities for each individuals in the network. By means of mathematical analysis, they prove the existence of an epidemic threshold below which the epidemic does not spread through the network. Using a different strategy, the article (NEPOMUCENO; TAKAHASHI; AGUIRRE, 2016) proposed and analyzed

a version of the IBM that uses exponential distributions to model the life expectancy and duration of the infectious period. The article describes in detail a framework that allows, among other things, to calculate the probability of eradication of a given disease.

This article proposes an IBM model based on (NEPOMUCENO; TAKAHASHI; AGUIRRE, 2016), which corresponds, on average, to the behavior of the SIR model, and extends this model to simulate the effect of two topologies of networks. The types of networks used in this work are:

- Random networks (IBM-RN): in which each node has the same probability of connecting to any other node in the network (ERDÖS; RÉNYI, 1959);
- Scale-free networks (IBM-SFN): in which few nodes have many connections and many nodes have few connections, ie the distribution of degrees follows a power law (ALBERT; BARABÁSI, 2002).

It performs a numerical computational case study that considers large scale networks implemented by an efficient framework. The implemented models are compared to each other and to the version of the IBM that corresponds, on average, to the continuous SIR model, in order to show the compatibility between them. Specifically, for IBM-RN, the effect of the variation in the average number of neighbors and the effect of the variation in the initial number of infected are studied. For IBM-SFN, the average temporal curve and the effect of the variation in the initial number of infected are analyzed. Furthermore, it compares the parameters, the basic reproduction number (R_0) and the mean of infected individuals in the final time (I^*) of the IBM-RN and IBM-SFN in relation to the equivalent SIR model, in a statistical point of view.

The paper is organized as follows. Section 2 describes the SIR and IBM models; Section 3 describes the IBM in Random and Scale-Free Networks; Section 4 shows and analyzes the obtained results, discussing the effects caused by the variation of the simulation parameters; and Section 5 concludes the paper and proposes future works.

2 SIR and IBM Models

This work considers the deterministic continuous SIR model and the IBM stochastic discrete model. In this section, the theoretical basis of epidemic propagation in SIR and IBM models is presented, as the equivalence between SIR and IBM.

2.1 SIR Model (Susceptible, Infected and Recovered)

The SIR model (Susceptible, Infected and Recovered) (KERMACK; MCKENDRICK, 1927) is one of the most used, since it represents the general classes and the basic mechanisms of the behavior of an epidemic (birth, mortality, infection, etc.). It uses the strategy of dividing the population into three compartments: Susceptible (S) representing individuals who can contract the disease if they come into

contact with another infected individual; Infected (I), representing those individuals who are carriers of the disease and can transmit it to other individuals; Recovered (R), representing those who have recovered from the disease and can not contract it again.

The model can be described by the system of ordinary differential equations:

$$\begin{cases} \frac{dS}{dt} = \mu N - \mu S - \frac{\beta}{N} IS, & S(0) = S_{t=0} \geq 0 \\ \frac{dI}{dt} = \frac{\beta}{N} IS - \gamma I - \mu I, & I(0) = I_{t=0} \geq 0 \\ \frac{dR}{dt} = \gamma I - \mu R, & R(0) = R_{t=0} \geq 0 \end{cases} \tag{1}$$

In which N is the population size; μ is the birth rate (new susceptible individuals), equal to the mortality rate; β is the transmission coefficient, that defines whether contagion occurs between susceptible and infected; and γ is the recovery rate. As the population size is set to N, it can then neglect the equation for $R(t)$, and by dividing by N the equations for susceptible and infected, we arrive at a percentage variables: $s(t) = S(t)/N$ and $i(t) = I(t)/N$.

Analyzing the dynamics of the SIR model, it can be determined the existence of two equilibrium points ($P1$ and $P2$). In $P1$, the population is free of disease and in $P2$, the population of infected reaches an endemic equilibrium:

$$\begin{cases} P_1(s_{f1}, i_{f1}) = (1, 0) \\ P_2(s_{f2}, i_{f2}) = \left(\frac{\mu+\gamma}{\beta}, \frac{\mu}{\mu+\gamma} - \frac{\mu}{\beta} \right) \end{cases} . \tag{2}$$

The stability of the equilibrium points can be analyzed by the concept of the basic reproduction number, R_0, which is the average of secondary infections produced when an infected individual is placed in a susceptible population (HETHCOTE, 2000; HEFFERNAN; SMITH; WAHL, 2005):

$$R_0 = \frac{\beta}{\gamma + \mu}. \tag{3}$$

If $R_0 \leq 1$ or $i_{t=0} = 0$, then the trajectories of the solutions will reach the eradication: P_1. If $R_0 > 1$, then the trajectories of the solutions with $i_{t=0} > 0$ will reach the endemic point P_1.

The parameters used in this work are: $\beta = 3.5$, $\gamma = 1/3$ e $\mu = 1/60$, which leads to $R_0 = 10$, which resembles the value shown by the AIDS epidemic in Kampala between 1985 and 1987 (ANDERSON; MAY, 1992). Additionally, P_1 is a saddle point (unstable) and P_2 is an asymptotically stable focus. Recently an analytical solution has been proposed for a similar SIR model (SHABBIR; KHAN; SADIQ, 2010). However, this work solves it numerically, using the Runge-Kutta's method, since the result would be used just in the comparison with IBM.

2.2 IBM (Individual-Based Model)

The Individual-Based Model (IBM) (KEELING; GRENFELL, 2000; SOLÉ et al., 1999; GRIMM, 1999; LOMNICKI, 1999; NEPOMUCENO; RESENDE; LACERDA, 2018) was proposed in order to evaluate computationally spreading of epidemics, in which each individual is represented by a group of unique and discrete characteristics. An IBM can be classified as a Cellular Automata (WOLFRAM, 1983), since individuals may be considered sites that may be in states. The states of each site are updated synchronously obeying probabilistic rules and the state of each site depends only on the state of its neighbors in the previous time step (a Markovian process).

Here, some premises based on the SIR model is used, such as (NEPOMUCENO; TAKAHASHI; AGUIRRE, 2016):

- Constant Population: size N;
- Categories of individuals: individuals are classified as Susceptible (S), Infected (I) and Recovered (R);
- Category changes: at each instant of time, an individual can change from one category to another (discrete-time transitions). The following changes may occur:
  - S, I, R $\rightarrow$ S: one individual dies and another one is born to take its place. If the individual does not die, the following changes may occur:
    - S $\rightarrow$ I: a susceptible makes contact with an infected and becomes infected;
    - I $\rightarrow$ R: an infected recovers from the disease and becomes recovered.
- Statistical distribution: all individuals have the same probability of contact; uniform distribution in each interval of time is defined for the mortality rate μ, the same of the birth rate, and for the recovery rate γ.

An individual $I_{l,t}$ is represented as follows:

$$I_{l,t} = [C_1 \quad C_2 \quad \dots \quad C_m], \tag{4}$$

where l is the individual in question, t is the instant of time, C_i is each characteristic that the individual presents and m is the number of characteristics. A whole population P_t is represented as follows:

$$P_t = [I_{1,t} \quad I_{2,t} \quad \dots \quad I_{n,t}]^T. \tag{5}$$

The initial population may be defined at random or may be set according to the purpose of the study. After evaluating all individuals, using a small Δt, the time is incremented by Δt. The algorithm is finalized when the time reaches its final value t_f.

Here, at each instant of time, each individual is evaluated, in which the following transitions may occur:

- each one is randomly chosen to death (and birth another susceptible) with probability $\mu\Delta t$;

- otherwise, if the individual is infected, another individual must be randomly chosen, who may be infected with probability $\frac{\beta\Delta t}{N-1}$;
- if the individual is infected, its recovery may be implemented with probability$\gamma\Delta t$.

2.3 Equivalence between SIR and IBM

This section justifies the proposed IBM, showing that its result converges, on average, to the result of the SIR model. For this, consider the following events, defined in the time interval $(k\Delta_t, (k+1)\Delta_t)$:

- $\Delta(k)$: one individual of the category I recovers, going to the category R;
- $\Omega(k)$: one individual dies and, consequently, another one is born in category S;
- $\Gamma(k)$: one individual of category S is randomly drawn;
- $\Phi(k)$: one individual has a contact with an individual of the category I;
- $\Lambda(k)$: one individual of the category S goes to the category I.

The probability of an individual to recover is: $P(\Delta(k)) = \gamma\Delta_t$, the probability of renewal of an individual is: $P(\Omega(k)) = \mu\Delta_t$, and the probability of new infections is: $P(\Lambda(k)) = P(\Gamma(k) \cap \Phi(k)) = P(\Gamma(k))P(\Phi(k))$, because they are independent events. Therefore, the probability of drawing a susceptible individual is:

$$P(\Gamma(k)) = \frac{S(k)}{N}, \tag{6}$$

the probability of contact with an infected individual is proportional to the number of infected persons present at that time:

$$P(\Phi(k)) = \frac{\beta\Delta_t}{N-1} I(k), \tag{7}$$

and, from Equations (6) and (7), follows that:

$$P(\Lambda(k)) = \frac{\beta\Delta_t}{N-1} \frac{S(k)I(k)}{N}. \tag{8}$$

The average number of individuals per unit of time of class I who recover from infection and pass into class R is given by: $\gamma\Delta_t I(k)$. The average numbers of individuals, per unit time, of the classes S, I and R that die and are replaced by others of the class S are given, respectively, by: $\mu\Delta_t S(k)$, $\mu\Delta_t I(k)$, $\mu\Delta_t R(k)$.

In IBM, at each time interval, the algorithm is executed once for each individual, that is, for N times. Thus, the average number of individuals, per unit of time,

going from class S to class I is given by:

$$N\frac{\beta\Delta_t}{N-1}\frac{S(k)I(k)}{N}. \tag{9}$$

For large values of N, Equation (9) can be rewritten as:

$$\frac{\beta\Delta_t S(k)I(k)}{N}. \tag{10}$$

These features allow to mathematically describe the dynamics of the average behavior of the IBM as:

$$\begin{aligned} S(k+1) = &\quad S(k) + \mu\Delta_t N - \mu\Delta_t S(k) - \frac{\beta\Delta_t I(k)S(k)}{N}, \\ I(k+1) = &\quad I(k) - \mu\Delta_t I(k) - \gamma\Delta_t I(k) + \frac{\beta\Delta_t I(k)S(k)}{N}, \\ R(k+1) = &\quad R(k) + \gamma\Delta_t I(k) - \mu\Delta_t R(k). \end{aligned} \tag{11}$$

Notice that Equation (11) matches with the discrete version of Equation (1), used to obtain solutions of the SIR model by numerical integration. Thus, it can be concluded that the proposed IBM model converges on average to the result of the SIR model.

3 IBM in Random and Scale-Free Networks

An important feature to consider in modeling epidemics is the social behavior of individuals. Generally, SIR and IBM models consider that all individuals are equally likely to relate to each other, ie contacts between individuals are randomly chosen following a uniform distribution. However, this is not found in today's societies, where we find individuals with different numbers and types of relationships.

Generally, in the epidemiological context, the nodes of the networks are the individuals of a certain population and the edges represent the contacts that can be made between two individuals.

The SIR and IBM models in their original formulation have the relationships between individuals represented by a complete graph, in which every node has edges connecting it with all others. In this work, we seek for models in which the relations (edges) between the individuals (nodes) of the population are represented by any network. The insertion of social networks in the SIR model is not a simple task, but in IBM it may be simpler, which makes this model a valuable tool. The types of networks used in this work are: random and scale-free.

3.1 Random Networks

The model of a random network has N nodes and k connections randomly chosen from the $N(N-1)/2$ possible connections (ERDÖS; RÉNYI, 1959). A

number of $C_N^k(N-1)/2$ equally likely different graphs can be generated. Therefore, the total number of connections is a random variable proportional to the probability p with expected value given by:

$$E(k) = p\left(\frac{N(N-1)}{2}\right) \tag{12}$$

The number of edges in each node follows a Poisson distribution, where histogram of the number of connections is shaped like a Gaussian. Therefore, these networks are also called Gaussian networks.

3.2 Scale-Free Networks

Some authors (ALBERT; BARABÁSI, 2002) have modeled a network that has a distribution of connections that follows a power law, ie few nodes establish many connections and many nodes establish few connections. Thus, the probability of a connection exists is given by:

$$P(k) = k^{-\gamma} \tag{13}$$

where k is the connectivity coefficient or number of connections and the exponent γ usually ranges between 2 and 3 for most real networks. As power laws are independent of any scale characteristic, these networks are now referred to as scale-free networks. Non-scale networks are present in a wide variety of real-world systems in several areas of knowledge.

The construction of a scale-free network can be done by adding new nodes to the current network in a progressive way. The network starts with N disconnected nodes. Firstly, a couple of nodes are randomly drawn, and the connection is made between them. After this initiation, an iterative process is executed in which, at each step, a node not yet connected is randomly drawn, and is connected to another randomly node already belonging to the network. In each addition, connections are established following the principle of preferential attachment, with probability described by:

$$P(k_i) = \frac{k_i}{\sum_{i=1}^{M} k_i} \tag{14}$$

where k_i is the number of connections of node i. When all nodes are already connected, it proceeds by drawing contacts between nodes belonging to the network, using the principle of preferential attachment, until a total of kN edges are reached.

3.3 Proposed IBM in Networks

Here, a new version of IBM is proposed, which considers that individuals are organized in any type of network (specifically, a complete graph, a random network or a scale-free network). This means a modification in the paradigm of the infection process: for each individual, look for the pair to be infected just in the list of adjacencies of the current individual, which corresponds to its neighbors.

Generally, the addition of different networks will greatly limit the spread of the epidemic. In fact, the spread of the disease will be slower, since using the same transmission coefficient (β), there are now far fewer neighbors to go through and thus possibly infect (YOUSSEF; SCOGLIO, 2011). The concept of density of the network can be used to find an adjusted β value, which behaves similarly to the original IBM. The value of the adjusted transmission coefficient for IBM with network is defined as $\bar{\beta}$ and the Equations (7) and (8) with the adjusted transmission coefficient $\bar{\beta}$ are changed by

$$P(\Phi(k)) = \frac{\bar{\beta} d \Delta_t}{N-1} I(k), \tag{15}$$

and, from Equations (6) and (7), follows that:

$$P(\Lambda(k)) = \frac{\bar{\beta} d \Delta_t}{N-1} \frac{S(k) I(k)}{N}. \tag{16}$$

which seeks to make the expansion of the epidemic in this network resemble the expansion in the original IBM using β, the Equation 11 can be rewrite, and the relationship between these two transmission coefficients is found:

$$\beta = \bar{\beta} d = \bar{\beta} \frac{k}{N-1} \tag{17}$$

where d is the density of the network (defined by $k/(N-1)$), k is the average number of contacts and N is the number of individuals in the population.

As in most of the research works in this area, static networks were used here, ie networks in which nodes and connections do not change over time. The explanation for this premise is that the dynamics of the epidemic is faster than the changes in the nodes of the network, because the average life span of individuals is much larger than the lifetime of an epidemic.

The proposed IBM has been implemented according to the Algorithm 1. The state changes occur with the probabilities described above, comparing with a number drawn with uniform distribution in the interval [0,1]. For death and recovery, it is to compare with $\mu \Delta t$ and $\gamma \Delta t$, respectively, and for infection it is to compare with $(\beta \Delta t)/(N-1)$ for each contact with another individual, which may be susceptible or not. Notice that this algorithm can be used for any kind of networks, including a complete graph, which would correspond to the original IBM.

4 Numerical Results

This section describes the results obtained through the simulations with the aim of achieving the objectives of this work: comparison between IBM and SIR models, effects of variation of some parameters of these models, behaviors observed after the addition of different networks to the IBM.

Simulations are performed in a population of $N = 490,000$ and average numbers of neighbors $k = 8$ and 24. The other parameters used are: $\Delta t = 0.1$ (time step),

Algorithm 1: IBM-networks

```
1  FirstInfected {Index of the initial infected}
2  Pop[:,0]← 0  {Generates an initial susceptible population}
3  Pop[FirstInfected,0]← 1 {Initial infection}
4  for t ← 1 to t_f step Δt do
5  |  for n ← 1 to N do
6  |  |  if rand() < μΔt then
7  |  |  |  Pop[n,t].estado ← 0 {Death}
8  |  |  else
9  |  |  |  if Pop[n,t−1].estado = 1 then
10 |  |  |  |  foreach i ∈ adjacency list of n do
11 |  |  |  |  |  if i ≠ n and rand() < βΔt/(N−1) then
12 |  |  |  |  |  |  if Pop[i,t−1].estado = 0 then
13 |  |  |  |  |  |  |  Pop[i,t].estado ← 1 {Infection}
14 |  |  |  |  |  |  end
15 |  |  |  |  |  end
16 |  |  |  |  end
17 |  |  |  |  if rand() < γΔt then
18 |  |  |  |  |  Pop[n,t] ← 2 {Recover}
19 |  |  |  |  end
20 |  |  |  end
21 |  |  end
22 |  end
23 end
```

$\mu = 1/60$, $\gamma = 1/3$, $\beta = 3.5$, $\bar{\beta} = 3.5(N-1)/k$, with the following initial conditions: $I_{t=0} = \{1,\ 1\%\ N,\ 10\%\ N\}$, $S_{t=0} = N - I_{t=0}$ and $R_{t=0} = 0$. The value of the infection coefficient $\bar{\beta}$ is defined according to Equation (17).

4.1 IBM in Random Networks (IBM-RN)

In this subsection, the effects caused by considering random networks in IBM (called as IBM-RN) are studied.

For first, the effect of the variation in the average number of neighbors 8 and 24 is studied. Figure 1 shows the average temporal curve of the IBM-RN by considering $k = 8$ and 24, in relation to the SIR curve. It is shown both the time evolution of infected in the initial stretch of the disease spread and in the final stretch, in order to highlight the differences among them. It is observed that as the number of neighbors increases, a larger number of infected persons is verified and this peak is reached more quickly. Furthermore, increasing the number of neighbors leads to a result closer to that obtained by the simulation of the SIR, both in the transitory and in the steady state. In fact, in the limit $k = N - 1$ case, the random network is

a fully connected network, which is the original IBM version, that has already been shown to be equivalent to the SIR model. It is worth noting that the final number of infected (I^*) grows as a function of the number of neighbors: SIR ($k = N - 1$), $k = 24$ and $k = 8$. However, the percentage difference is small (between 3.5% and 4.3%). Considering 8 neighbors, it is possible to verify that R_0 presents an average value of 4.93. Using $I_{t=0} = 1$ the eradication of the disease was verified in only 6 of the 100 executions. Using $I_{t=0} = 1\%\ N$ and $I_{t=0} = 10\%\ N$ there was a similar behavior regarding the equilibrium, but with no cases of eradication of the disease. Considering 24 neighbors, the number of eradications for the values of $I_{t=0}$ was exactly the same, but R_0 presented a larger value, corresponding to an average value of 6.55.

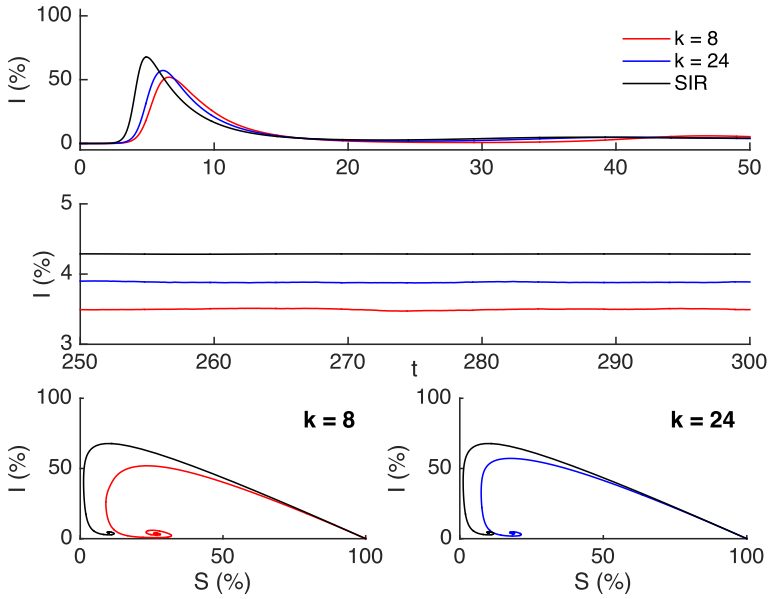

Figura 1 - Mean temporal evolution curves in IBM-RN varying the number of neighbors. Temporal evolution of the infected population with a focus on the initial stretch of the disease spread and in the final stretch; and the phase plans for each of the values of k. The parameters used for the simulation of SIR and IBM are: $N = 490,000$, $\Delta t = 0.1$, $k = \{8, 24\}$, $\mu = 1/60$, $\gamma = 1/3$, $\beta = 3.5$, $\bar{\beta} = \{214375.0, 71458.1875\}$, $S_{t=0} = 489999$, $I_{t=0} = 1$ and $R_{t=0} = 0$.

Now, the effect of the variation in the initial number of infected is analyzed. Figure 2 shows the effect of this variation on the temporal intensity of infected and susceptible people in a random network of 8 neighbors, in relation to SIR. Just like

in SIR, and therefore in the IBM original form, as $I_{t=0}$ increases there is an increase in the peak intensity, which occurs more rapidly. As the peak becomes smaller and later, the curve becomes smoother, which does not occur in the SIR model. It is believed that this smoothing is due to the lower number of initially infected individuals that makes the spread of disease slower in the random network. This same result is obtained for different numbers of neighbors, as in the case $k = 24$, wherein as the number of neighbors increases the result is getting closer to that obtained in the SIR model.

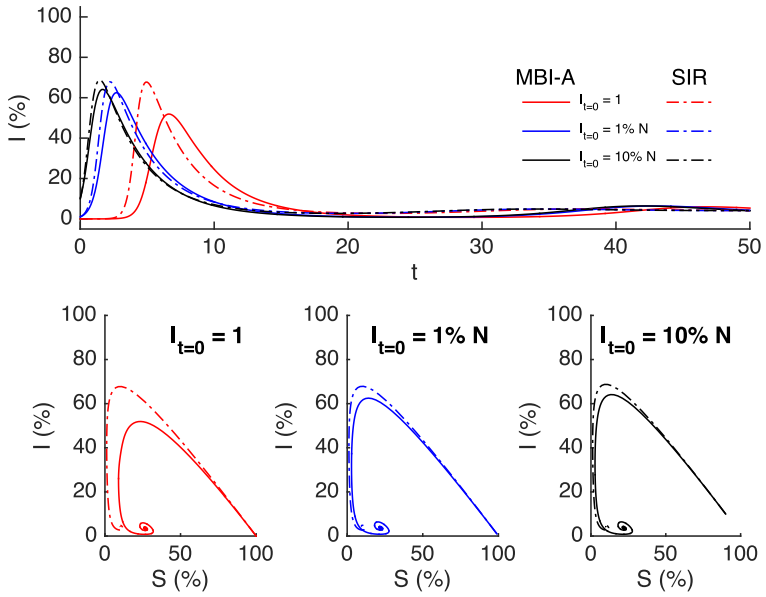

Figura 2 - Mean temporal evolution curves in IBM-RN varying the initial number of infected. Temporal evolution of the infected population and the phase plans for each of the values of $I_{t=0}$. The parameters used for the simulation of SIR and IBM are: $N = 490000$, $\Delta t = 0.1$, $k = 8$, $\mu = 1/60$, $\gamma = 1/3$, $\beta = 3.5$, $\bar{\beta} = 214375.0$, $I_{t=0} = \{1,\ 1\%\ N,\ 10\%\ N\}$, $S_{t=0} = N - I_{t=0}$ and $R_{t=0} = 0$.

The impact of the variation of the initial number of infected in the final number of infected is analyzed using $k = 8$ neighbors. Figure 3 shows the boxplot of the values of I^* for each initial infected: 1, 1% N e 10% N, and the value observed in the SIR model. The 6 eradication cases that occurred in the case of just 1 initially infected were withdrawn. It can be seen that the values of I^* of the samples of IBM-RN are inferior to the SIR, even though a much higher transmission coefficient ($\bar{\beta}$) was used. From the Lilliefors test, each sample of IBM-RN was found to be

normal at 95% confidence; by the Bartlett test, they are homocedastic with 95% confidence. The data are independent as they were generated using an appropriate random number generator. Since the premises of ANOVA have been respected, it shows that there are no statistical evidences, with 95% confidence, that there are differences between the means of I^* varying the initial number of infected, just as it occurs in the SIR model and IBM in their original formulation, which indicates that the initial number of infected influences in the transient state but not in the stationary state. With fewer infected individuals initially, the disease expands more slowly, but goes to the same equilibrium.

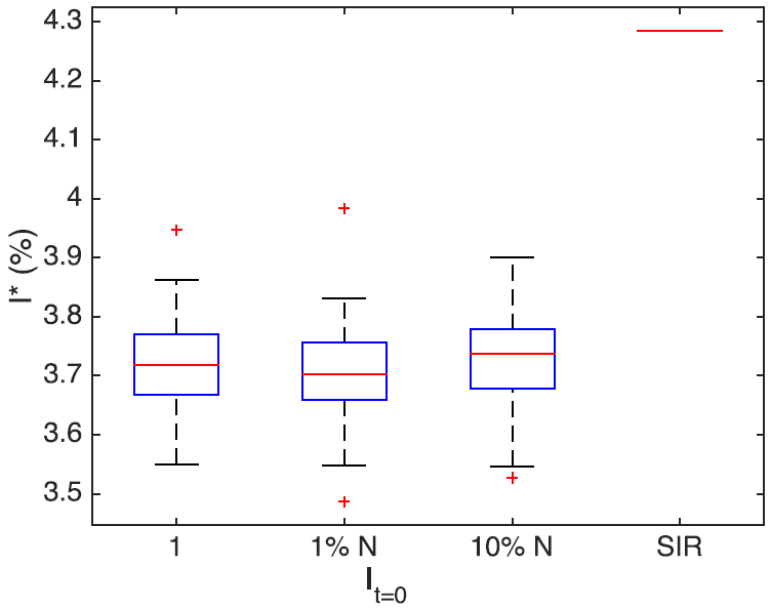

Figura 3 - Boxplots of the final infected values I^* for different initial number of infected $I_{t=0}$ in IBM-RN.

It is believed that the proposed IBM-RN may better represent a real population, as an alternative to traditional SIR/IBM models, making possible to occur individuals with different neighbors. However, since the degree distribution of individuals is described by a Poisson distribution, the most individuals will have near-average numbers of neighbors, which may not approach to the reality.

4.2 IBM in Scale-Free Networks (IBM-SFN)

In this subsection, the effects caused by the addition of the scale-free network to IBM (called as IBM-SFN) are studied, considering an average number of neighbors fixed equal to $k = 8$.

Figure 4 shows the average temporal curve of the IBM-SFN, in relation to the SIR curve. It is shown both the time evolution of infected in the initial stretch of the disease spread and in the final stretch, in order to highlight the differences among them. It can be observed that in the scale-free network, unlike the random network, the rise to the infected peak starts faster than in the SIR model, even if the peak value is lower. This indicates that the network topology causes the disease to spread more rapidly at the outset, even though the same number of individuals are not contaminated, than in IBM in its original formulation. Here again, it is possible to infer that the final number of infected I^* grows as a function of the number of neighbors, with the SIR value as a limiting factor, even though the percentage is small, with values ranging approximately between 3% and 4.3%. For only 1 initially infected individual, it is possible to verify that R_0 presented an average value of 3.94. The disease was eradicated just in 8 out of 100 trials. In simulations using $I_{t=0} = 1\%\ N$ and $I_{t=0} = 10\%\ N$ similar behavior was observed in equilibrium and no cases of disease eradication occurred.

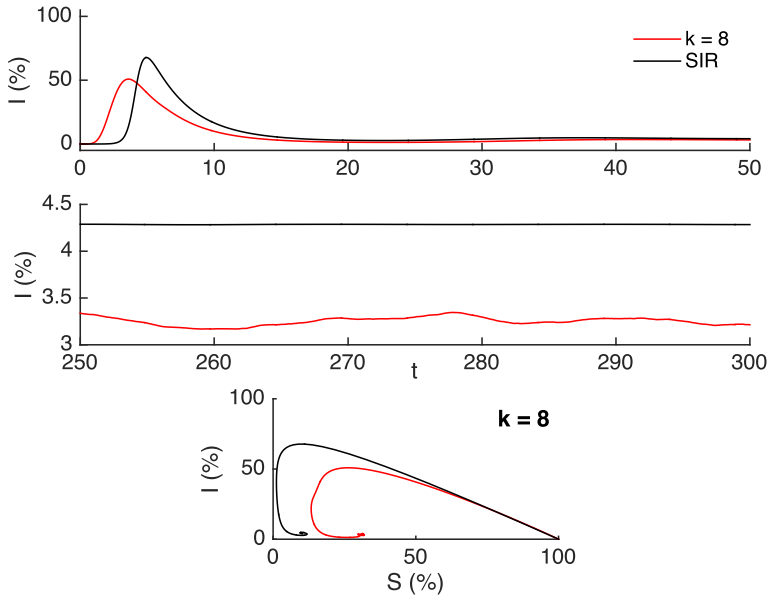

Figura 4 - Mean temporal evolution curves in IBM-SFN for the number of neighbors $k = 8$. Temporal evolution of the infected population with a focus on the initial stretch of the disease spread and in the final stretch; and the phase plans. The parameters used for the simulation of SIR and IBM are: $N = 490,000$, $\Delta t = 0.1$, $k = 8$, $\mu = 1/60$, $\gamma = 1/3$, $\beta = 3.5$, $\bar{\beta} = 214375.0$, $S_{t=0} = 489999$, $I_{t=0} = 1$ and $R_{t=0} = 0$.

Now, the effect of the variation in the initial number of infected is analyzed. Figure 5 shows the effect of this variation on the temporal intensity of infected and susceptible people in a scale-free network of 8 neighbors, in relation to SIR. For IBM-SFN with $k = 24$ a similar behavior is observed. Just like in IBM-RN, as $I_{t=0}$ increases there is an increase in the peak intensity, which occurs more rapidly, and as the peak becomes smaller and later, the curve becomes smoother.

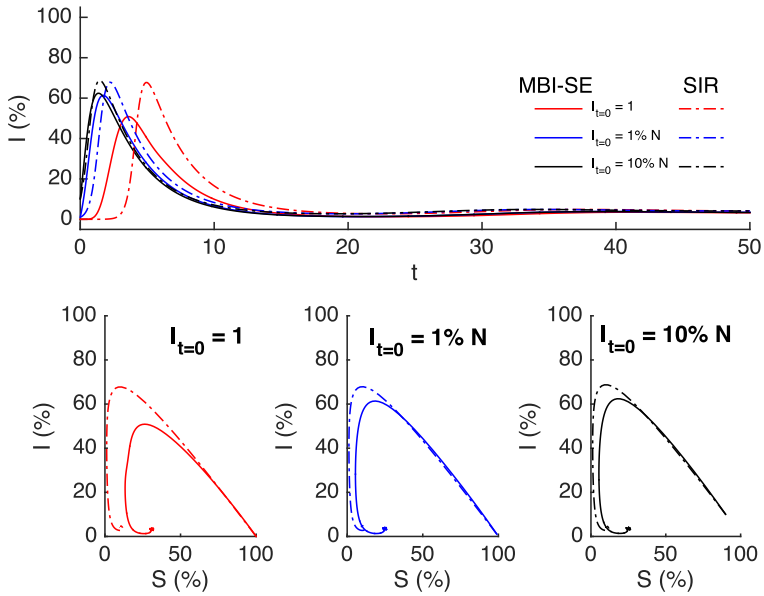

Figura 5 - Mean temporal evolution curves in IBM-SFN varying the initial number of infected. Temporal evolution of the infected population and the phase plans for each of the values of $I_{t=0}$. The parameters used for the simulation of SIR and IBM are: $N = 490,000$, $\Delta t = 0.1$, $k = 8$, $\mu = 1/60$, $\gamma = 1/3$, $\beta = 3.5$, $\bar{\beta} = 214375.0$, $S_{t=0} = 489999$, $I_{t=0} = \{1,\ 1\%\ N,\ 10\%\ N\}$ and $R_{t=0} = 0$.

The impact of the variation of the initial number of infected 1, 1% N e 10% N in the final number of infected is analyzed. Figure 6 shows the boxplot of the values of I^* for each initial infected, and the value observed in the SIR model. The 8 eradication cases that occurred in the case of just 1 initially infected were withdrawn. As in IBM-RN, it can be seen that the values of I^* of the samples are inferior to the SIR, even though a much higher transmission coefficient ($\bar{\beta}$) was used. Thus, the expansion of the disease in the scale-free network is faster than in SIR, although the infected peak is smaller. It is also verified that with fewer infected individuals at the beginning, the disease expands more slowly, but goes to the same equilibrium of cases

where there were more initially infected individuals, just as occurred in the random network. Using the Lilliefors test with 95% confidence, it can be seen that premise of normality of ANOVA has not been respected. For this reason, the Kruskal Wallis test was performed, which shows that there are no statistical evidences, with 95% confidence, to affirm that there are differences between the means of I^* observed when the initial number of infected in IMB-SFN with 8 neighbors, just as it occurs in the SIR, IBM in its original formulation and in IBM-RN. It is noteworthy that, by the Levene test, with 95% confidence, the samples are homocedastic and the data are independent since they were generated through the use of an appropriate random number generator.

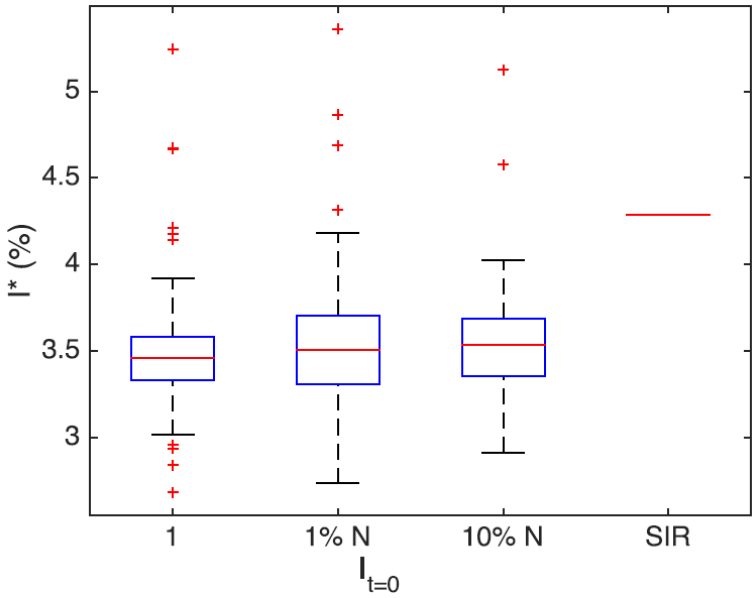

Figura 6 - Boxplots of the final infected values I^* for different initial number of infected $I_{t=0}$ in IBM-SFN. The parameters used for the simulation of SIR and IBM are: $I_{t=0} = \{1,\ 1\%\ N,\ 10\%\ N\}$, $S_{t=0} = N - I_{t=0}$ and $R_{t=0} = 0$.

It is believed that the scale-free network can better represent a real population since it is based on a construction in which random contacts are drawn, taking into account the criterion of preferential attainment. It is verified in its distribution of the neighbors with the occurrence of many individuals with few contacts and few individuals with many contacts. It is believed that this is approaching reality since there are people who are actually more popular and therefore more connected while others are less connected.

4.3 Comparison between IBM in Random and Scale-Free Networks

The experiment of this subsection compares the parameters of the IBM model in random (IBM-RN) and scale-free (IBM-SFN) networks with the parameters of the equivalent SIR model, showing if it is possible to adjust the time evolution of these IBM, so that they are statistically equivalent to the respective trajectories of the SIR model, as can be done for the original IBM. Specifically, in relation to the SIR parameters (γ, $\bar{\beta}$ and μ) and to the IBM estimated parameters ($\hat{\gamma}$, $\hat{\beta}$ and $\hat{\mu}$), the following null hypotheses are defined:

$$\begin{aligned} H_0^{\gamma}: &\quad \hat{\gamma} = \gamma \\ H_0^{\beta}: &\quad \hat{\beta} = \overline{\beta} \\ H_0^{\mu}: &\quad \hat{\mu} = \mu \ . \end{aligned}$$

Figure 7 show the relative frequency histograms for the IBM estimated parameters, within the respective Gaussians with same means and standard deviations, their theoretical value and the 5% critical values of significance of the tests. In this simulation, the IBM estimated parameters are:

- IBM-RN: $\hat{\gamma} = 0.0517$, $\hat{\beta} = 0.1056$ and $\hat{\mu} = 0.0104$;
- IBM-SFN: $\hat{\gamma} = 0.0673$, $\hat{\beta} = 0.1504$ and $\hat{\mu} = 0.0117$.

Therefore, considering IBM-RN, the test does not reject none of the hypotheses H_0^{γ}, H_0^{β} and H_0^{μ} with a significance level of 5%. In relation to IBM-SFN, the hypothesis H_0^{μ} is not rejected, but hypotheses H_0^{γ} and H_0^{β} is rejected with level significance of 5%. This means that the average IBM-RN result can be obtained by simulating an appropriate SIR model, which should not occur for the average IBM-SFN result.

Figure 8 illustrate the infected and susceptible curves obtained with the mean of the IBM evolutions and the curves of the corresponding SIR model. It can be highlighted that, despite the reasonable adjustment in the number of infected of the SIR model in relation to the IBM mean curves, the number of susceptible curve has notable differences in IBM-SFN. Hence, a study is made on the susceptible curves, in order to validate this difficulty in the adjustment of the parameters of the SIR model for this network. Indeed, for each IBM susceptible curve, the residue of the curve is defined as the sum of the residues at each point of the curve:

$$R(S) = \sum S - S_{SIR}. \tag{18}$$

Considering the same parameters estimated from the SIR model, Monte Carlo simulations are made using IBM in its original formulation to generate the null hypothesis of the mean of the residuals. The residue of IBM in networks and of IBM in its original formulation are denoted by R_R and R_M, respectively. The alternative hypotheses are considered in order to verify the differences found in the susceptible curves. The level of significance used in all tests is $\alpha = 0.05$. Table 1 shows the results obtained from this analysis. Notice that, in fact, the empirical p-values found corroborate with the susceptible graphs of Figure 8.

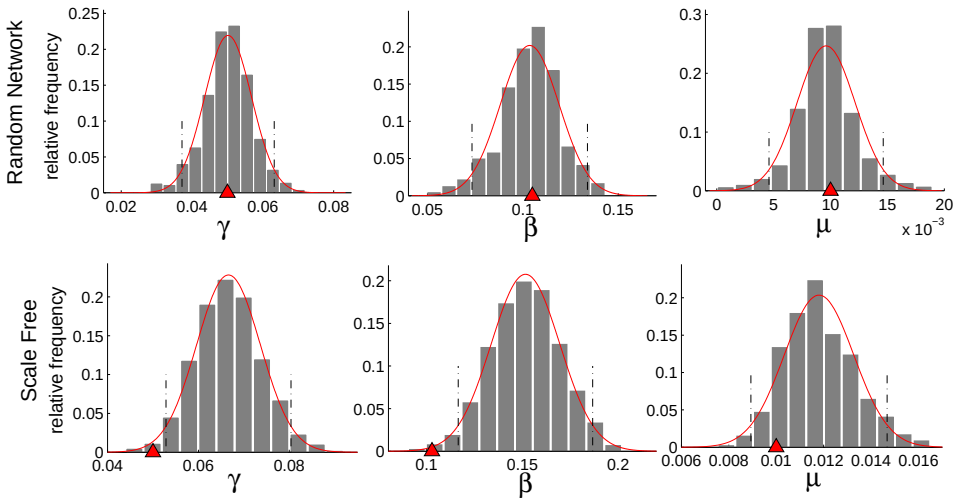

Figura 7 - Histograms of the parameters estimated for IBM-RN and IBM-SFN. Gaussian curves of the same means and standard deviations are also shown. The parameters used for the simulation of SIR and IBM are: $N = 490,000$, $\Delta t = 0.1$, $k = 8$, $\mu = 1/60$, $\gamma = 1/3$, $\beta = 3.5$, $\overline{\beta} = 214375.0$, $S_{t=0} = 489999$, $I_{t=0} = 1\%$ and $R_{t=0} = 0$.

IBM in networks	$\mathbf{H_0}$	$\mathbf{H_1}$	**empirical p-value**
IBM-RN	$R_R = R_M$	$R_R \neq R_M$	0.85
IBM-SFN	$R_R = R_M$	$R_R > R_M$	0.10

Tabela 1 - Statistical analysis of the number of susceptibility curves obtained by IBM in networks. The p-values indicate the percentage of IBM executions in networks that lead not to reject the null hypothesis H_0, unfavorable to the alternative hypothesis H_1.

Therefore, it can be concluded that the inhomogeneity of the distribution of connections between the nodes differentiates the models with respect to their dynamic behavior. If a community is organized according to a complex random network, then the IBM of this network behaves according to the SIR model, with a new transmission rate $\overline{\beta}$, but statistical analyzes indicate that if a community is organized according to a complex scale-free network, the IBM behavior of this network has difficulties in being modeled by the SIR model, with a new transmission rate that preserves its meaning.

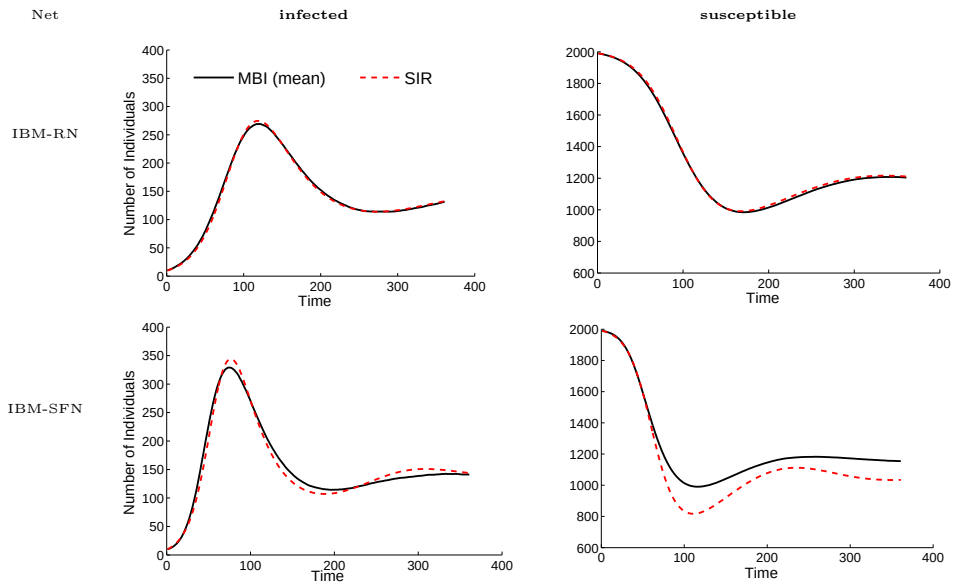

Figura 8 - Comparison between the mean infected and susceptible curves of IBM-RN and IBM-SFN in relation to the SIR model. The parameters used for the simulation of SIR and IBM are: $N = 490,000$, $\Delta t = 0.1$, $k = 8$, $\mu = 1/60$, $\gamma = 1/3$, $\beta = 3.5$, $\bar{\beta} = 214375.0$, $S_{t=0} = 489999$, $I_{t=0} = 1\%$ and $R_{t=0} = 0$.

Finally, in order to compare the networks with respect to the basic reproduction number (R_0) a group of 100 simulations of each model with the same parameters as before, but now with the initial condition of only one individual initially infected and all others susceptible.

Table 2 summarizes the data found and it can be verified that the value of the mean R_0 is higher in the random network than in the non-scale. This was already expected in the non-scale network because of the existence of many individuals with few neighbors and few with many ones. In the random network, most individuals have the same number of neighbors, which means that the disease spreads more easily. It is also verified that the value presented for the SIR has always been higher than the one observed in the other models. It can also be observed that the higher the value of R_0, the greater the value of the mean infected individuals in the final time I^*.

5 Conclusions

This paper proposed a version of IBM that is proved to converge, on average, to the result of the SIR model, and studied the behavior of this IBM in two different

Model	$\mathbf{R_0}$	$\mathbf{I^*}$
IBM-RN	4.93	3.494
IBM-SFN	3.94	3.214
SIR	10	4.285

Tabela 2 - Comparison of the mean values of basic reproduction number R_0 and the number of infected individuals in the final time I^* in the models IBM-RN, IBM-SFN and SIR.

types of network: random and scale-free networks, seeking to better represent what is expected to happen in a more realistic population. In fact, the use of different networks showed different behaviors in the spread of an epidemic in relation to the original IBM:

- The peak of infected in IBM in networks is always lower than in the SIR model and the rise and fall is smoother. While in IBM-RA the peak of infected is always later than in SIR, the opposite occurs in IBM-RSE. By increasing the average number of neighbors, the curve of infected individuals is closer than in the SIR model.
- The dynamics present differences throughout the transient regime, being that the scale-free network has a faster expansion than in the SIR model, which is faster than in the random network. However, the means of the final infected do not change in the models in networks. This indicates that the network topology does not seem to influence the permanent state, only the transient state.
- The mean value of R_0 is lower in the scale-free network, then in the random network, and are both always smaller than the theoretical value of the SIR model. The increase in the mean value of R_0 was reflected in the increase in the mean value of the final number of infected.
- The scale-free IBM does not seem to be tuned to the SIR model, as can occur in the original IBM and in IBM in random networks. It was concluded that the non-homogeneity of the distribution of links between the nodes distances the models with respect to their dynamic behavior, having the SIR model as a baseline.

Other approaches may be undertaken in future works, such as the implementation of mechanisms to control the epidemic by using of vaccination or isolation.

Acknowledgments

The authors thank to FAPEMIG and CNPq for their financial support. Also, the authors thank reviewers and editors for their comments.

SILVA, C. R. R.; ALMEIDA, A. C. L.; CARDOSO, R. T. N.; TAKAHASHI, R. H. C. Modelos Epidêmicos Baseados em Indivíduos Aplicados em Redes Aleatórias e sem Escala. *Rev. Bras. Biom.,* Lavras, v.38, n.1, p.102-124, 2020.

- *RESUMO: Este trabalho propõe uma versão do Modelo Baseado em Indivíduos (MBI) que converge, em média, para o resultado do modelo SIR (Susceptible-Infected-Recovered), e estuda o efeito deste MBI em dois tipos de redes: aleatória e sem escala. Um estudo de caso computacional numérico é realizado, utilizando redes de larga escala implementadas por uma estrutura eficiente. Testes estatísticos são realizados para mostrar as semelhanças e diferenças entre os modelos de redes e o modelo determinístico tomado como referência. Resultados de simulação verificam que diferentes topologias de redes alteram o comportamento da propagação da epidemia nos seguintes aspectos: evolução temporal, reprodutibilidade basal e número de infectados no final.*

- *PALAVRAS-CHAVE: Modelo Baseado em indivíduos; redes aleatórias; redes sem escala.*

Referências

AŁAR, S. et al. Agent based modelling and simulation tools: A review of the state-of-art software. *Computer Science Review*, v. 24, p. 13–33, 2017.

ALBERT, R.; BARABÁSI, A.-L. Statistical mechanics of complex networks. *Reviews of Modern Physics*, v. 74, n. 1, p. 47, 2002.

ANDERSON, R. M.; MAY, R. M. *Infectious diseases of humans: dynamics and control*. New York: Oxford University Press, 1992.

BADHAM, J.; STOCKER, R. The impact of network clustering and assortativity on epidemic behaviour. *Theoretical Population Biology*, v. 77, n. 1, p. 71–75, 2010.

BARTHELEMY, M. et al. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Physical Review Letters*, v. 92, n. 17, p. 178701, 2004.

EAMES, K. T. D. Modelling disease spread through random and regular contacts in clustered populations. *Theoretical Population Biology*, v. 73, n. 1, p. 104–111, 2008.

ERDÖS, P.; RÉNYI, A. On random graphs I. *Publicationes Mathematicae*, Institute of Mathematics, University of Debrecen, Hungary, v. 6, p. 290–297, 1959.

GANESH, A.; MASSOULIÉ, L.; TOWSLEY, D. The effect of network topology on the spread of epidemics. In: *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.* [S.l.: s.n.], 2005. v. 2, p. 1455–1466.

GE, Y. et al. A synthetic computational environment: To control the spread of respiratory infections in a virtual university. *Physica A: Statistical Mechanics and its Applications*, v. 492, p. 93–104, 2018.

GRIMM, V. Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecological modelling*, v. 115, n. 2, p. 129–148, 1999.

HE, Z.; MIEGHEM, P. V. The spreading time in sis epidemics on networks. *Physica A: Statistical Mechanics and its Applications*, v. 494, p. 317–330, 2018.

HEFFERNAN, J.; SMITH, R.; WAHL, L. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, v. 2, n. 4, p. 281–293, 2005.

HETHCOTE, H. W. The mathematics of infectious diseases. *SIAM review*, v. 42, n. 4, p. 599–653, 2000.

HWANG, D. et al. Thresholds for epidemic outbreaks in finite scale-free networks. *Mathematical Biosciences and Engineering*, v. 2, n. 2, p. 317–327, 2005.

KEELING, M. J.; GRENFELL, B. T. Individual-based perspectives on R_0. *Journal of Theoretical Biology*, v. 203, n. 1, p. 51–61, 2000.

KERMACK, W. O.; MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, v. 115, n. 772, p. 700–721, 1927.

KISS, I. Z. et al. Mathematics of epidemics on networks. *Cham: Springer*, 2017.

LOMNICKI, A. Individual-based models and the individual-based approach to population ecology. *Ecological modelling*, v. 115, n. 2, p. 191–198, 1999.

MACAL, C. M.; NORTH, M. J. Tutorial on agent-based modelling and simulation. *Journal of simulation*, v. 4, n. 3, p. 151–162, 2010.

MAGAL, P.; RUAN, S. Susceptible-infectious-recovered models revisited: From the individual level to the population level. *Mathematical biosciences*, v. 250, p. 26–40, 2014.

MILLER, J. C. Spread of infectious disease through clustered populations. *Journal of the Royal Society Interface*, v. 6, n. 41, p. 1121–1134, 2009.

NEPOMUCENO, E.; RESENDE, D.; LACERDA, M. A survey of the individual-based model applied in biomedical and epidemiology. *Journal of Biomedical Research Review*, v. 1, n. 1, p. 11–24, 2018.

NEPOMUCENO, E. G. et al. Individual-based modelling and control of bovine brucellosis. *Royal Society open science*, v. 5, n. 5, p. 180200, 2018.

NEPOMUCENO, E. G.; TAKAHASHI, R. H. C.; AGUIRRE, L. A. Individual based-model (ibm): An alternative framework for epidemiological compartment models. *Revista Brasileira de Biometria*, v. 34, p. 133–162, 2016.

NEWMAN, M. E. The structure and function of complex networks. *SIAM review*, v. 45, n. 2, p. 167–256, 2003.

NEWMAN, M. E. J. Random Graphs with Clustering. *Physical Review Letters*, v. 103, n. 5, p. 058701, 2009.

NOEL, P.-A. et al. Time evolution of epidemic disease on finite and infinite networks. *Physical Review E*, v. 79, n. 2, p. 026101, 2009.

NOWZARI, C.; PRECIADO, V. M.; PAPPAS, G. J. Analysis and control of epidemics: A survey of spreading processes on complex networks. *ArXiv e-prints*, 2015.

PASTOR-SATORRAS, R.; VESPIGNANI, A. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, v. 63, n. 6, p. 066117, 2001.

PASTOR-SATORRAS, R.; VESPIGNANI, A. Epidemic spreading in scale-free networks. *Physical Review Letters*, v. 86, n. 14, p. 3200–3203, 2001.

PASTOR-SATORRAS, R.; VESPIGNANI, A. Immunization of complex networks. *Physical Review E*, v. 65, n. 3, p. 036104, 2002.

RAKOWSKI, F. et al. Influenza epidemic spread simulation for poland?a large scale, individual based model study. *Physica A: Statistical Mechanics and its Applications*, v. 389, n. 16, p. 3149–3165, 2010.

SERRANO, M. A.; BOGUNA, M. Percolation and epidemic thresholds in clustered networks. *Physical Review Letters*, v. 97, n. 8, p. 088701, 2006.

SHABBIR, G.; KHAN, H.; SADIQ, M. A note on exact solution of SIR and SIS epidemic models. *arXiv preprint arXiv:1012.5035*, 2010.

SOLÉ, R. V. et al. Controlling chaos in ecology: from deterministic to individual-based models. *Bulletin of Mathematical Biology*, Springer, v. 61, n. 6, p. 1187–1207, 1999.

WOLFRAM, S. Statistical mechanics of cellular automata. *Reviews of Modern Physics*, v. 55, n. 3, p. 600–644, 1983.

YOUSSEF, M.; SCOGLIO, C. An individual-based approach to sir epidemics in contact networks. *Journal of Theoretical Biology*, v. 283, n. 1, p. 136–144, 2011.

Received on 21.01.2019.

Approved after revised on 06.08.2019.