Computer Methods and Programs in Biomedicine 197 (2020) 105715

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Viral disease spreading in grouped population

Tomasz Gwizdałła

Faculty of Physics and Applied Informatics, University of Łódź, Pomorska 149/153, Łódź 90-236, Poland

ARTICLE INFO

Article history:
Received 16 April 2020
Accepted 15 August 2020

Keywords:
Agent-based modeling
COVID-19
Disease spreading
Influenza
SEIR model
Scale-free networks

ABSTRACT

Background and Objective: The currently active COVID-19 pandemic has increased, among others, public interest in the computational techniques enabling the study of disease-spreading processes. Thus far, numerous approaches have been used to study the development of epidemics, with special attention paid to the identification of crucial elements that can strengthen or weaken the dynamics of the process. The main thread of this research is associated with the use of the ordinary differential equations method. There also exist several approaches based on the analysis of flows in the Cellular Automata (CA) approach.

Methods: In this paper, we propose a new approach to disease-spread modeling. We start by creating a network that reproduces contacts between individuals in a community. This assumption makes the presented model significantly different from the ones currently dominant in the field. It also changes the approach to the act of infection. Usually, some parameters that describe the rate of new infections by taking into account those infected in the previous time slot are considered. With our model, we can individualize this process, considering each contact individually.

Results: The typical output from calculations of a similar type are epidemic curves. In our model, except of presenting the average curves, we show the deviations or ranges for particular results obtained in different simulation runs, which usually lead to significantly different results. This observation is the effect of the probabilistic character of the infection process, which can impact, in different runs, individuals with different significance to the community. We can also easily present the effects of different types of intervention. The effects are studied for different methods used to create the graph representing a community, which can correspond to different social bonds.

Conclusions: We see the potential usefulness of the proposition in the detailed study of epidemic development for specific environments and communities. The ease of entering new parameters enables the analysis of several specific scenarios for different contagious diseases.

1. Introduction

The recent pandemic related to the worldwide expansion of the COVID-19 coronavirus redirected the attention of scientists and entire societies to techniques that can be helpful when analyzing and predicting the spread of diseases in communities. Viruses are always present in our environment, and this presence is usually uninteresting. For example, we are so familiar with most influenza virus mutations that the epidemics caused by them are widely considered mainly in economic discussions. Only occasionally, when some particularly aggressive mutations emerge or during pandemics caused by some coronaviruses (SARS, MERS), is the public alerted about the potential danger.

With the knowledge about possible directions of disease transfer and the number of people at risk, the identification of the most vulnerable groups is crucial. In 1927, Kermack and MacKendrick published their paper [1], which is now considered the first approach to the mathematical modeling of epidemic processes. They introduced the crucial concepts by classifying members of a population, when considering an illness, as susceptible and infected and tried to find the relationships between these classes with the help of differential calculus. This division is also used today, and the number of applications using, based on their approach, ordinary differential equations (ODE) continues to increase. We want to emphasize a topic that is difficult to address when using the ODE approach - the inclusion of stochastic effects. Indeed, there are different models in which some external force, either a periodic, e.g., seasonality [2], or purely stochastic [3], here for the SIS

E-mail address: tomasz.gwizdalla@uni.lodz.pl

https://doi.org/10.1016/j.cmpb.2020.105715
0169-2607/

model, force, assuming the susceptibility of an individual recovering from a disease, is used.

In our approach, we follow another approach that is related to the analysis of the distribution of particular groups in real human communities. This approach is often related to the Cellular Automata model. The typical approach can be found, e.g., in Sirakoulis' paper [4]. In the model, the population is distributed over a two-dimensional space, and this space is, as is typical for CA, divided into squares. The spread of disease is modeled by the deterministic rules describing the transitions between states. Many investigations use similar models, e.g., the famous model proposed by Ferguson's group [5] conducts much more soft division to generate the areas. We pay special attention to this model since it was later widely used to study the possibilities of preventing disease. In [6,7], the expected results of different intervention procedures during the influenza epidemic in Asia and the United States, respectively, are shown. The same model has recently been applied to the COVID-19 [8] pandemic. Great Britain and the United States were the areas of application, but according to authors' remarks, every high-income country can be studied with this model.

A similar approach using CA can also be observed in [9], where the transfer between cells is the main factor supporting the spread of disease. The exchange of people is also a crucial feature of the model presented by Holko [10]. The results are here reproduced for the whole area of Poland, divided into 36×36 squares, showing the possible spread of the influenza epidemic. In our earlier paper [11], we showed, taking into account the same area of inhabitance as in this paper, that the model requires special attention, primarily due to the existence of so-called size effects.

Although similar to the approaches based on cellular automata, our approach is different, mainly due to the change in the definition of the topology of interpersonal links. We do not consider the aggregate number of individuals in a particular state and a particular area, but we create direct links between them. Thus, we can regard our model as an agent-based model. After defining the topology and the set of features characterizing agents, we can individualize their behavior. In the paper, we use the Barabasi-Albert (BA) model of the creation of a community graph [12,13]. This approach currently seems to be the most popular among the numerous attempts to model communities, dated from the seminal paper on random graphs by Erdos and Renyi [14]. The main property of the BA model is that it leads to a power-law distribution of nodes according to their degree (the scale-free property), which is typical of communities. Therefore, several real-world networks can be described by a BA model, e.g., a world wide web network, an actor or scientific collaboration, or even the E. Coli metabolism [15,16]. Several authors have already proposed the use of the BA network [17–20], but their models concentrate on other problems. For simplicity, we do not introduce the differences in the behavior of agents.

There are also papers where some of the concepts used in our paper were considered. Ramos [21] studied the case of grouping on some form of a two-dimensional grid. Balcan [22] studied the role of hubs, well known from complex network theory, in outbreak prevention by their identification and vaccination. Hellewell [23] studied the graph of direct links, where a negative binomial distribution determined the probability of the creation of links.

Among the papers related to the current COVID-19 pandemic, we want to pay attention especially to these documents that can be used in our calculations or give some additional ideas for disease analysis. We mention here the reports concerning the incubation time [24–26] and the ones that present information about important parameters, such as the basic reproduction number R_0 [27,28]. Boldog's paper [27] also draws interesting conclusions concerning the potential risks for particular countries when taking into account their connections with China.

This paper is organized as follows: In the next section, we present a model emphasizing two revealed problems - the construction of a graph describing a society and the procedure of the transfer of illness. We also justify the idea of division into groups based on some, generally arbitrary, factor. In the section devoted to the presentation of results, we concentrate on the epidemic curves, which are presented in two forms, i.e., the number of new cases and the number of recovered persons (in the absence of the mortality rate), and on the analysis of intervention, considered as the minimization of the number of contacts between neighbors in the network. We also show the effect of using different procedures on the possibility of people in different groups becoming sick.

2. Model

The most popular way to model the disease-spreading process is the method of Kermack [1], which is based on individuals being assigned to one of several groups. In their original paper, the authors did not use the contemporary terminology, but later, the acronym SIR began to be used.

In the SIR model, the members of a population can be classified as belonging to one of three groups concerning a disease: susceptible (S) - those who can become ill; infective (I) - those who can infect others; and recovered (R), those who are permanently immune. The set of ordinary differential Eq. (1) determines the number (or fraction) of individuals in a particular state and at a particular time.

$$\begin{aligned} \frac{dS(t)}{dt} &= -aS(t)I(t) \\ \frac{dI(t)}{dt} &= aS(t)I(t) - bI(t) \\ \frac{dR(t)}{dt} &= bI(t) \end{aligned} \tag{1}$$

In the above formula, S, I, and R are the numbers of people in the respective phases of illness, a is the contact rate, and b is the inverse of the infectious period.

This approach was later extended by including a fourth phase - exposed (E) - between phases S and I. This fourth phase describes the fraction of the population in the latent (incubation) phase of an illness. This addition enables us to include more realistic processes when considering the majority of infectious diseases. The set of ODEs now takes the form:

$$\begin{aligned} \frac{dS(t)}{dt} &= \mu(N(t) - S(t)) - \beta\frac{I(t)}{N(t)}S \\ \frac{dE(t)}{dt} &= \beta\frac{I(t)}{N(t)}S - (\mu + \delta)E(t) \\ \frac{dI(t)}{dt} &= \delta E(t) - (\mu + \gamma)I(t) \\ \frac{dR(t)}{dt} &= \gamma I(t) - \mu R(t) \end{aligned} \tag{2}$$

Some parameters in formula 2 are the same as in 1: β is the contact rate, and γ is the inverse of the infectious period. Additionally, we have to consider $N(t)$ - the total number of people in a community, satisfying the condition $N(t) = S(t) + E(t) + I(t) + R(t) = const$ - and some new parameters, such as δ, which is the inverse of the latent period, and μ, i.e., the mortality (and also birth) rate. The equality of mortality and birth rates is certainly the condition enabling us to consider a constant $N(t)$. We should emphasize that both of the above sets of equations are exemplary since their detailed forms depend strongly on the system's assumptions, but they present the idea of the calculations. The presented models also offer many opportunities to be modified or expanded. We can mention here, e.g., SEIS, where no immunity is assumed after a disease is passed on, or SEIJR, where an additional

J phase (isolated individuals with treatment) is placed between the I and R phases.

This continuous approach enables easy calculation of one of the most interesting values describing the potential effect of an outbreak, i.e., the basic reproduction number, which is a simple function of the ODE parameters:

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

2.1. Modeling the community graph

Although continuous models based on the ODEs give many interesting and practical results, it is well known [3] that there exists a large stochastic effect in the epidemic process. Therefore, we propose considering epidemics in a more individual manner. We start from two assumptions:

- every infection is a result of the interaction between infected and susceptible,
- we have to find the network of interpersonal connections initially and then model the spread of an illness.

The second of these two points distinguishes our proposition from some of the abovementioned approaches, being based mainly on the Cellular Automata methodology, where the particular cells in a two-dimensional lattice correspond to different sections, covering the whole real area under investigation, for example, certain parts of cities or countries (see, e.g., [10,20]). We instead numerically follow a scheme in which the possibility of every individual interaction is considered separately. The fundamental idea of the BA model is the observation regarding the growth of the network. Instead of creating links between the existing nodes, the preferential character of this growth is assumed, which governs the creation of a graph corresponding to a community network. In conclusion, the individuals are added to the existing graph one after the other, and during this process, the straightforward formula:

$$P(k) = \frac{k_i}{\sum k_j} \tag{4}$$

is used when determining the probability of linking a new node to an existing node indexed by *i*. k_i here is the degree of node *i*, and the summation in the denominator runs over all existing nodes.

We present the results for the two cases of graph modeling. The first one is the pure BA model, as it is well known from the seminal papers. However, we propose a modified approach, which is related to the fact that every individual is a member of different groups. As a group, we understand here different subsets of a community described by some common interest or features. Good examples include the place of residence or the workplace. Even if we do not consider someone connected in the sense of a typical social network, we can more easily meet him/her in a local store or the company hallway. This is why we decided to modify the probability given by formula 4, trying to take into account the mentioned effects. The solution that imposes itself is to change the relative probabilities of acceptance for links connecting nodes/individuals belonging to the same group when compared to those from other groups. However, no clear evidence exists regarding how we can introduce these changes. It could strongly depend on several additional details, for example, an individual living in a block of flats and shopping at the supermarket can accidentally meet many more unknown people than a resident of a cottage in the suburbs, who shops only at a local store. Since we have to aggregate all these possibilities, we decided to use the mechanism where the probability of connecting nodes is tripled for nodes belonging to the same group, while for different groups, it is divided by 2.

$$P_{same\,group}(k) = min(1,\ 3 * \frac{k_i}{\sum k_j})$$
$$P_{different\,groups}(k) = 0.5 * \frac{k_i}{\sum k_j} \tag{5}$$

This procedure has a different significance in different phases of the creation of a graph of connections. In the early phase, the values of probabilities (4) are relatively large; thus, almost all nodes in the same group are connected (probability is close to 1), and the decreased probability for different groups starts to play an important role just after the initialization. The later added nodes usually have smaller degrees; hence, both formulas 5 have comparable significance. The choice of particular multipliers (3 and 0.5) in Eq. (5) has no particular background. We want to clearly distinguish between the case where belonging to some environment strongly influences an individual's ability to be connected with another person and that where no particular preferences are possible. In this paper, we use three types of graphs describing communities:

- the pure BA graph;
- a graph based on assignment to one of four groups with completely artificial sizes;
- a graph based on assignment to one of 16 groups with sizes based on the inhabitance of particular areas of a selected medium-sized city in Poland.

We create the four groups mentioned in the second of the above points to introduce the visible differences in their sizes. The idea is to observe the effects of the proposed division in the simplified case, and the percentages of people in successive groups are {0.1, 0.2, 0.3, 0.4}. The second division is based on a real analysis of our city - Łódź (Lodz). Easily accessible information [29] shows that we can approximate the city's symmetric shape as a square with an edge length close to 16 km. The number of inhabitants in Lodz can be estimated to be approximately 700000. The large square is further divided into 16 smaller squares (4 rows by 4 columns). For this approach, we can establish the number of inhabitants in particular parts of the city. The results of this division are schematically presented in Fig. 1 (a). The shades correspond to the number of inhabitants in a particular area, and the numbers inside the squares are the populations of groups in thousands.

To show in more detail the distribution of people in the studied case, we also added Fig. 1 (b), where a similar division is shown for the system of $16 \times 16 = 256$ smaller groups.

It is a fundamental assumption that this community is isolated; thus, we do not have the chance for potential secondary outbreaks, although it is not difficult to introduce this effect.

The modification of the probability distribution in creating the connection is shown in Fig. 2. We show the distributions for the three types of graphs mentioned above when creating them for 70000 individuals. We can observe that the algorithm creates less than 10^{-4} of pairs with a probability greater than 0.01. The differences are visible, but it may seem that they are not able to produce a significant effect on the spreading process. There is also no significant difference between distributions for different divisions (4 groups or 16 groups).

2.2. Modeling the disease transfer

The process of disease transfer is, in reality, difficult to describe in the language of mathematical formulas. When considering the direct contact results, we have to take into account many details, such as the duration of contact or the distance between individuals. Indeed, there are also many features of medical origin, such as

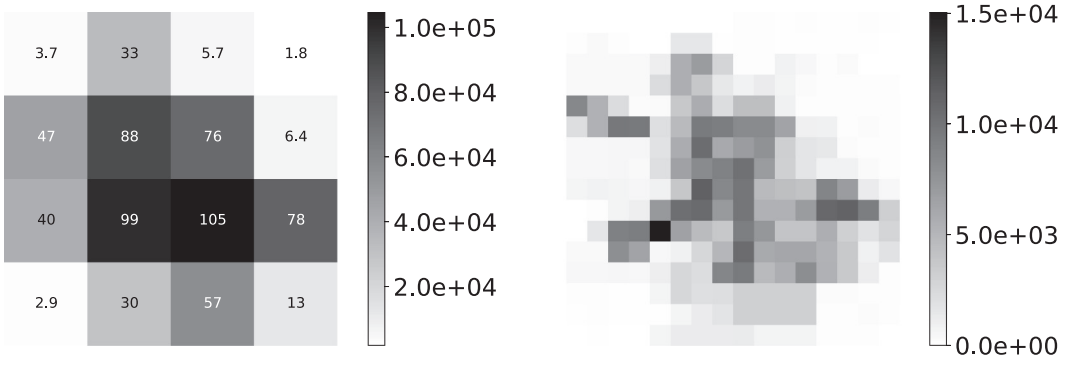

Fig. 1. The number of inhabitants of Lodz, corresponding to the number of people in particular groups when dividing into 16 (a) or 256 (b) groups.

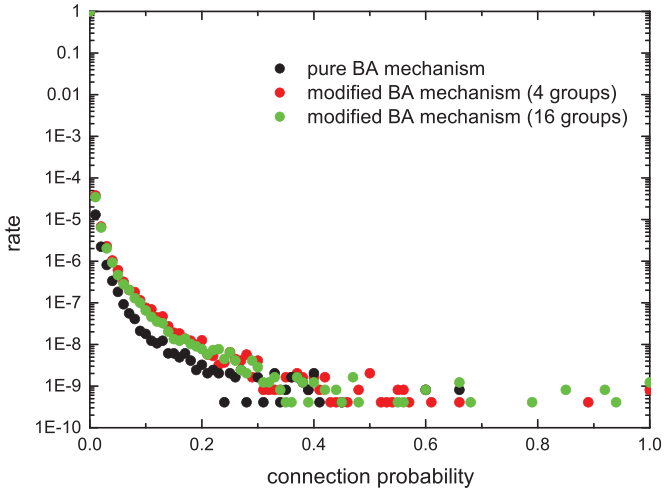

Fig. 2. Comparison of the connection probability distributions

the individual immune system or blood group [30]. Due to these problems, a much more simplified approach is typically used. For example, in Schimit's paper [31], the probability of infection depends on the number of infected neighbors (ν): $p(\nu) = (1 - e^{-k\nu})$. In our approach, we consider the process of possible infection separately during every step and for every possible pair of neighbors.

There are two probabilities that are the basis of our model:

- p_M - probability of meeting (contact),
- p_I - probability of infection.

The first value (p_M) is related to the social reasons for infection - without contact with a pathogen, we cannot become ill. Therefore, we have to define the value that describes the probability that two individuals can meet during the time corresponding to one time step. Thus, p_M is the value that describes the possibility of a situation when contact is long enough to enable the transmission of illness. The second value (p_I) is related to the medical-related processes of infection and should, in general, correspond to the characteristics of the pathogen causing the disease. In the simplification, to meet the current paper's needs, this value is the probability of transferring an illness during a meeting, which occurs with probability p_M. In general, this value can undoubtedly correspond to many processes related to the pathogen transfer between two individuals. As a result, this method enables distinguishing the influence of processes of different origin on the result of simulation. By changing p_M, we can introduce, e.g., the effect of a quarantine, and by changing p_I, we can distinguish different diseases. It is very important to mention here that p_I does not change during the simulation run. This change would be a good model for the reaction of a pathogen to the environmental conditions, such as the temperature, humidity, or UV radiation level.

Table 1
The list of parameters used in the modeling process.

Number	Parameter	Acceptable set of values	
		Influenza	COVID-19
1	Size of sample	{7000, 70000}	
2	Mode of graph creation	{pure BA, modified BA}	
3	Probability of meeting	0.1	
4	Probability of infection	0.25	
5	Number of days in exposed (E) state	2	0
6	Number of days in infected (I) state	4	14
7	Number of groups	{4, 16}	
8	Index of group where outbreak starts	{0, 3, 10}	
9	Choice of patient 0	{the hub in the group}	
10	Day of intervention	{no intervention, 10}	

The crucial problem when studying the expansion of an epidemic is to parametrize the SEIR model by introducing into the calculations the realistic times at which an individual stays in the Exposed (t_E) and Infectious (t_I) states. In this paper, we show the results for two pairs of times, making it possible to distinguish between the two types of infections. As a first case, we choose the times typical for influenza ($t_E = 2$, $t_I = 4$), following the cases studied in [10,32]. For the second case, we choose times that best resemble the COVID-19 data. Since there exist some preliminary data [25,26,33], we decide to set ($t_E = 0$, $t_I = 14$).

Finally, the scheme of disease transfer is as follows:

- We start from exactly one "patient 0".
- The location of this patient varies between simulation runs, but it is always unambiguously defined. Usually, "patient 0" is the hub of the selected group, and sometimes it is a less connected node. The detailed information about him/her is always given when describing a particular result.
- One time step is equivalent to one day.
- During the day, an individual is in group I - infectious - if he/she can transfer the disease to every one of his/her connected individuals with total probability $p_T = p_M * p_I$.
- After the infection time, every individual stays recovered (R) and does not return to the susceptible group (S).
- The mortality and birth rates are assumed to be 0.

In Table 1, we present a summary of the simulation parameters. Three fields in the table need further explanation. Unless stated otherwise, the outbreak starts in the so-called hub in the most populous group. In the concept of social networks, hubs are those nodes that are characterized by the highest degree. This means that the selection of the initial point of the outbreak is not random but is assigned to the node with a relatively high possibility of spreading the disease.

We also decide to propose the model of intervention in the form of the separation of individuals. The main idea of prevention forced by authorities during the COVID-19 outbreak is to keep people at home. Certainly, for most of the population, it is not possible to completely resign from leaving places of isolation; hence, they significantly decrease the number and intensity of outside contacts. We model this by decreasing the p_M factor. If we force an intervention, every day after its start, p_M is halved ($p_{M,t+1} = 0.5 * p_{M,t}$ - t enumerates the time steps) until it reaches a value lower by one order of magnitude compared to the starting value $p_M = 0.1$. For the initial calculations, we assume that intervention, if it occurs, starts on the 10th day.

3. Results

In this section, we present the results of calculations made for the model presented above. We start the analysis from the information on the size of graphs used in particular simulations. Two values are chosen, i.e., 7000 and 70000, with several reasons justifying these choices. First, we can expect the existence of a size effect. The size effect is typical for dynamical system simulations and manifests itself in the dependence of results on the size of the sample. We have observed this effect for disease spreading [11] models, and we can also expect its existence for these calculations. The choice of particular values is done to achieve easy scaling and comparison with the number of nodes in the subsequent sections of Fig. 1. All the values presented there are determined on the basis of the number of Lodz inhabitants, which is approximately 700000; thus, the real number of 3700 persons in the upper-left section corresponds to 37 or 370 for 7000 and 70000, respectively. A smaller number of nodes can also make the calculations faster and, therefore, enable collecting more extensive statistics. The most time-consuming process is the creation of the network - its time complexity is $O(n^2)$.

In Fig. 3 and 4, we show the epidemic curves and the cumulative epidemic curves for selected cases described by the parameters listed in Table 1. In Fig. 3, we show the daily number of new cases. In Fig. 4, as cumulative data, we consider all individuals who passed the infectious state (I). For simplicity, we call them recovered in the figure. The results presented in the plots are averaged over ten runs for every set of parameters. This procedure makes all curves smoother, and all statistical effects disappear; we will return to them later. We prepare both figures in the same style. Having the data for two illnesses and two cases related to reducing the contact probability (denoted as a form of intervention), we show them in such a way that the upper plots correspond to the influenza-related data and the lower ones to the COVID-19 related data. The left plots show the results without any intervention, while the right ones correspond to the inclusion of the social distancing effect by decreasing the p_M parameter.

The pure epidemic curves (Fig. 3) show that every factor included in the simulation parameters can influence the course of epidemics. Since we show percentages, we can directly compare all curves. The most visible differences can be observed for the influenza-related data. As one can expect, when starting the disease in the hub of the most populated group (described as a hub), we obtain the highest rate of total infected persons. This result is, however, obtained for the particular division into four groups. With this division, the increase in connection probability plays such a significant role that it causes a large increase in the number of persons with illness. Interestingly, the same case but considered for the pure BA model or for a larger number of groups (16) does not cause this significant effect. It can also be observed that the maximum of epidemics takes place earlier than for the corresponding larger number of groups for the smaller groups. This effect is visible for all pairs of curves prepared for the same set of parameters but different sizes. It is also essential that the difference in maximum percentages for different parameters can differ by approximately an order of magnitude, and epidemics can last a long time. Indeed, it never expires, as is known for influenza, certainly with different intensities.

When looking at the plots for influenza with intervention, we can estimate the progress of influenza when we assume that the restrictions, similar to those introduced during COVID-19, are implemented on the 10th day from the time of outbreak. We can expect that in approximately three weeks (10 days after implementing social distancing), the disease will disappear. Indeed, we can consider this situation as a thought experiment, since no government would impose restrictions due to the seasonal flu.

We can compare our flu-related observations from both figures (3 and 4). The second figure enables us to easily compare the total numbers of people passing on the flu. As one can see, the final number of recovered persons and the outbreak's dynamics vary with the choice of model. When trying to compare the results with the real-world data, we can use, for example, the information provided by the Polish National Institute of Hygiene. The data for Poland, with just under 40 million citizens, show that, for different years, the percentage of influenza cases is estimated to be between 2 and 15 percent (see http://www.old.pzh.gov.pl/oldpage/epimeld/grypa/Ryc_3.jpg). We observe that our size-dependent results are in the appropriate range.

When looking at the COVID-19-related data in the lower plot of Fig. 3, we notice that the differences here are visibly less important. As we observed earlier, the maxima of the curves for smaller sizes occur slightly earlier than for the corresponding curves for larger sizes. This difference is, however, visibly smaller. The longer infectious time causes the disease to be more aggressive and faster. For almost all cases, after approximately three months, the disease disappears. The number of people spreading the disease can be seen in Fig. 4. The final number of ill persons reaches a value in the interval [57%, 73%]. This result is certainly unacceptable, especially taking into account the fact that the real mortality index for COVID-19 is reported to be approximately 5%.

The influence of intervention on the tenth day is shown in Fig. 3 (d). The crucial observation is that the disease distinctly decreases its intensity after several dozens of days. This profile is similar to the curves obtained, for example, in some Western European countries (see, e.g., https://www.worldometers.info/coronavirus for the Netherlands or Spain). The disease stays significantly weaker after approximately two months from the first infection. Indeed, these systems are not closed ones.

We can show that the size effect is present in this particular type of calculation. Except for the data for COVID-19 parameters without intervention, every other case strongly depends on community size. The most straightforward way to notice this effect is to observe the curves drawn with the same line type but different colors. This means that the spread of disease depends strongly not only on the relative ratio of connections inside and outside groups but also on the absolute number of links. Some extended calculations may be needed, especially to describe the influenza-type epidemic.

The interesting effect that we can observe is the change in characteristics with the increase in the number of groups. We can see that the total number of ill persons for the division into four groups is usually higher than for the pure BA model and later decreases when dividing the community into 16 groups. In Fig. 4, this relation is observed, e.g., when comparing the plots for $size = 7000$, initialized in the hub of a large group, and 3 models described as BA, 4gr, and 16gr. We think that the reason for this property is the fact that we strongly support the creation of links inside groups (see Eq. (5)). Thus, if there is a small number of relatively numerous groups, the average number of links could

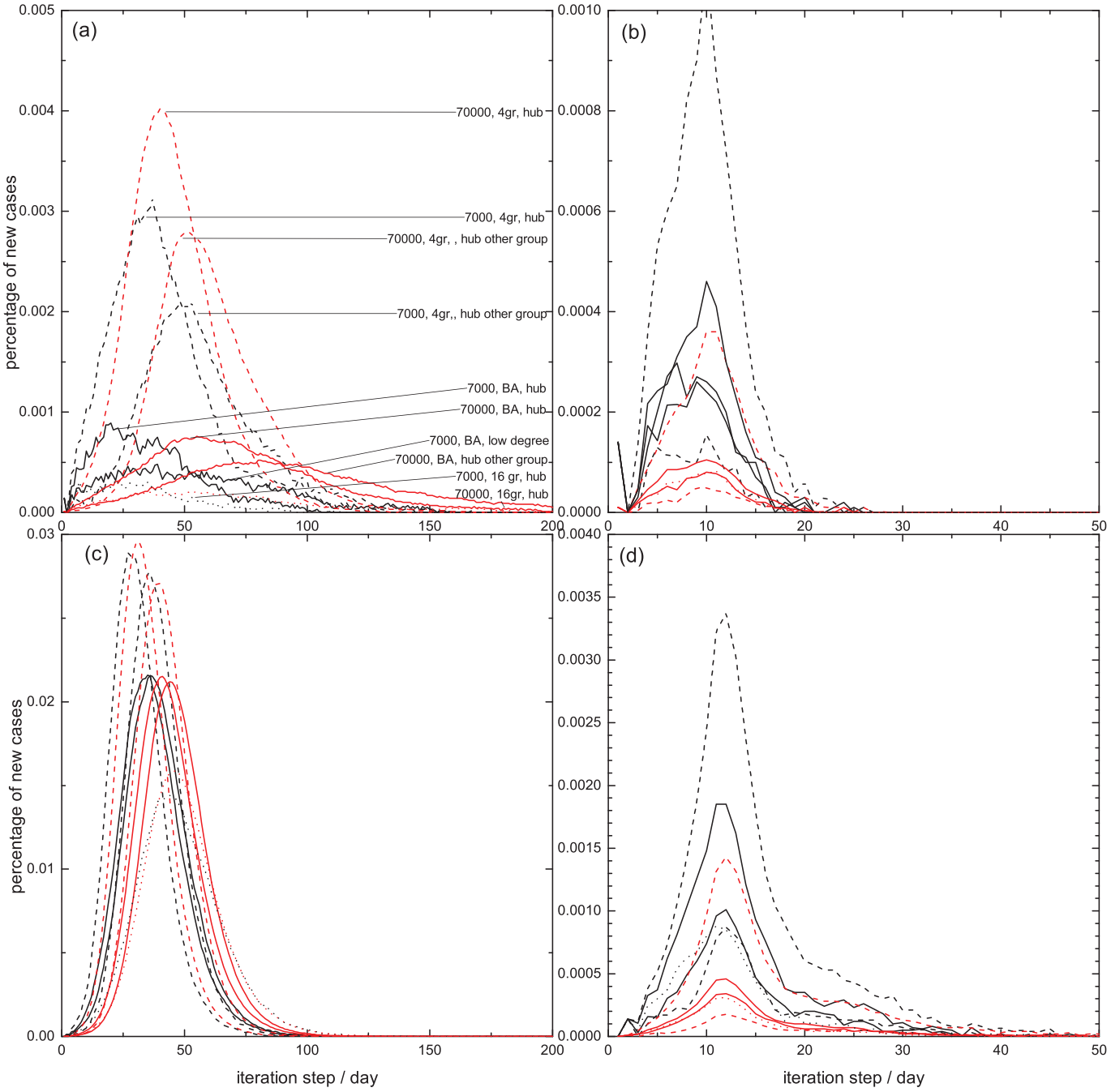

Fig. 3. The course of epidemic curves (new cases) for the sets of parameters selected among those described in Table 1. The image is organized as follows: Among the four plots, the two upper plots (a and b) show the results for the influenza epidemic ($t_E = 2$, $t_I = 4$); the two lower plots (c and d), for COVID-19 ($t_E = 0$, $t_I = 14$); the two left plots (a and c) correspond to the situation without any intervention; and the two right plots (b and d), to the intervention of social type with, as described in the text, the limitation of contacts by the change in the parameter p_M. Every plot contains 10 curves for different scenarios and conditions of modeling. The color of the line corresponds to the size of the sample: black presents data for 7000 nodes in a graph, while red, for 70000 nodes. The style of the curve distinguishes the method used for graph creation: solid line - pure BA model (BA), dashed line - 4 groups (4gr), and dotted line - 16 groups (16gr). We also show some plots describing when patient 0 is the one with the greatest number of links (hub) in the most populous group and some others where the hub is in the less populated group. The abbreviations in parentheses correspond to the description of curves shown in the first plot.

increase, strengthening the effect of disease transfer. We should emphasize the magnitude of these changes, which can make some results even ten times greater than others.

We also tested the significance of the choice of the first infected individual. To prevent the plot from being unreadable, we limited the number of different cases to just two. For the first one, for which most of the calculations are made, “patient 0” is in the most populous group. For three selected individuals, described as “hub other group”, he/she is also a hub but located in the less populated group. The difference in the final number of ill persons can be, when analyzing this feature, approximately up to 30% higher than when starting from the more populated subgroup.

Considering once more the effect of intervention (see plots (b) and (d) in Fig. 3 and 4), we can observe that, with intervention in-

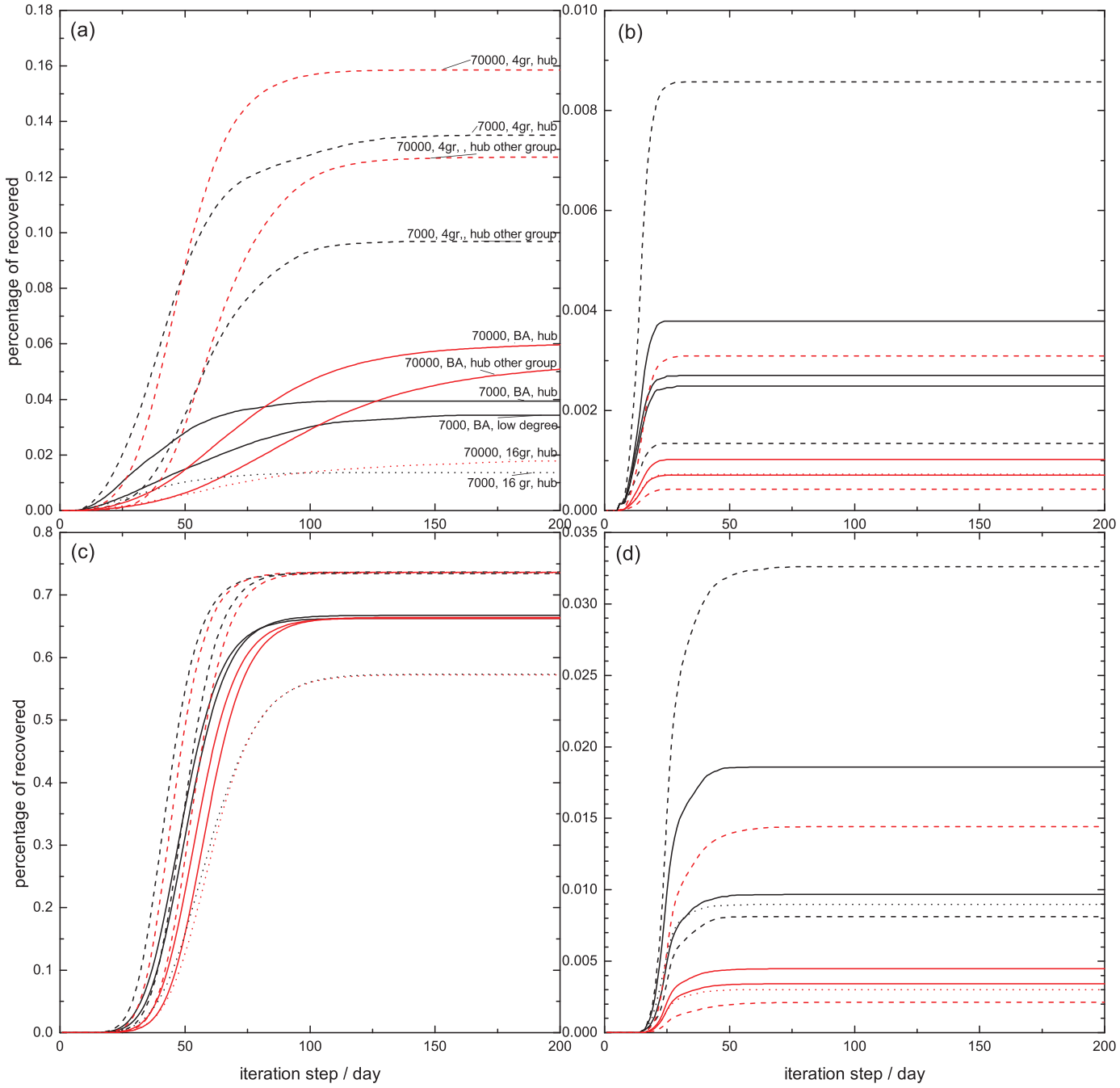

Fig. 4. The course of cumulative epidemic curves for the same sets of parameters as in Fig. 3. The organization of the image is the same as in Fig. 3.

cluded, the duration of the epidemic does not strongly depend on the parameters of the model. Some dispersion certainly exists but is not significant. To show some characteristics, we sampled the selected averaged values for particular simulation types in Table 2. We choose to present three points in time:

- the time when the cumulative curve changes its character from convex to concave: $t_{inflection}$
- the time when the cumulative function reaches 0.99 of its maximal value: $t_{0.99}$
- the time when the cumulative function reaches 0.999 of its maximal value: $t_{0.999}$

For all types of simulation (a-d), the data are averaged over the whole set of parameters, and averages (along with standard deviations) are shown. The results show that for the virus with the longer infectious period, the duration is shorter (but certainly causes many more cases) than for those with shorter times t_E and t_I. We must pay attention to the fact that the results here are strongly correlated with the model used. The size effect is observable and characterized by a stronger influence of intervention for larger samples, and the division into 16 groups prevents the spread of disease. These data also give us information about the final phase of an epidemic. The most important result is shown in plot (d). Even with a fast, radical, and widely accepted mechanism of intervention, we can still expect that after ten weeks, approx-

Table 2
The averaged data of the characteristic times in days (time steps).

Type of simulation	$t_{inflection}$	$t_{0.99}$	$t_{0.999}$
Influenza without intervention (plot a)	49.8 ± 15	144.7 ± 36	160.7 ± 36
Influenza with intervention (plot b)	13.6 ± 1	24.8 ± 2	28.2 ± 3
COVID-19 without intervention (plot c)	49.8 ± 5.3	92.7 ± 10	113.2 ± 12
COVID-19 with intervention (plot d)	23.8 ± 0.4	53.8 ± 7.6	69.8 ± 8

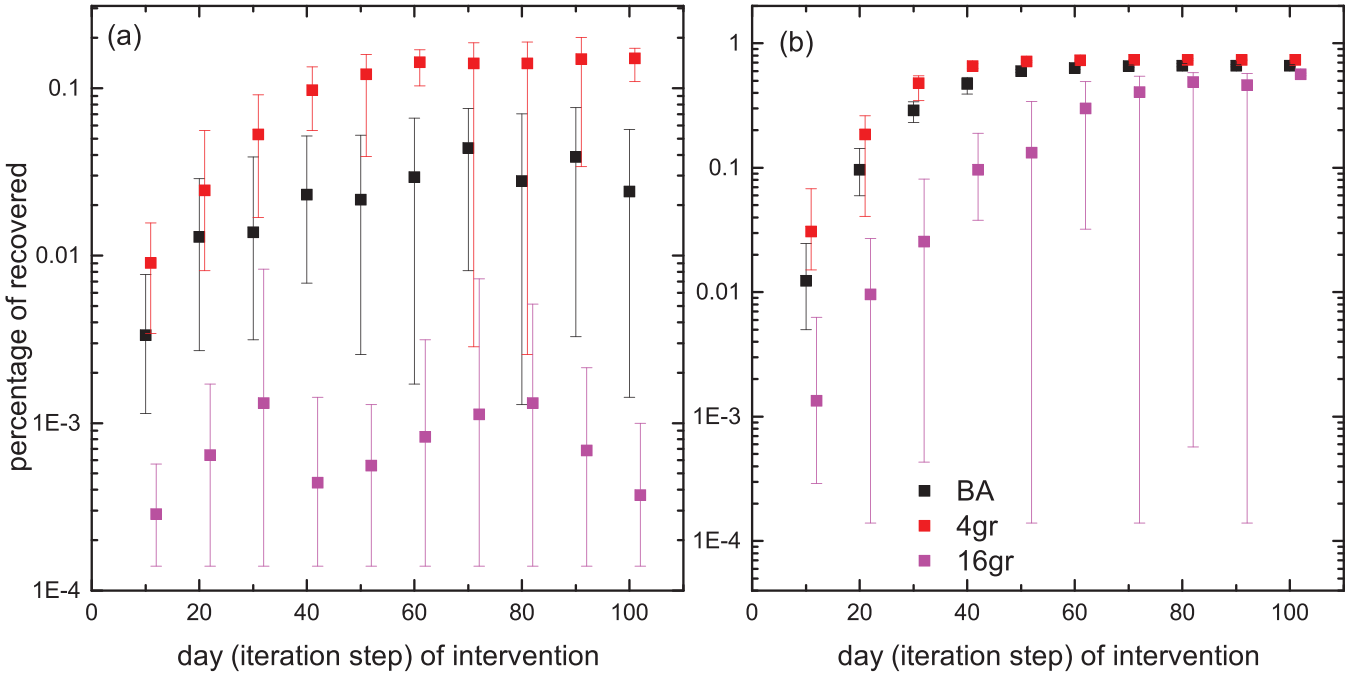

Fig. 5. The effect of the delay of intervention. Plot (a) is prepared for data used for influenza ($t_E = 2$, $t_I = 4$); plot (b), for COVID-19-related parameters ($t_E = 0$, $t_I = 14$). We perform the calculations for a sample size equal to 7000 and with the start of the outbreak in the hub of the most populous group. Then, we prepare plots for three cases: pure BA model, four groups, and 16 groups. The descriptions are the same as in the earlier figure, i.e., BA, 4gr, and 16gr, respectively. The bars correspond to the range obtained during different runs for particular parameters.

imately 1 person in every 100000 (1 per mil due to $t_{0.999}$ being approximately 1 percent the total number of ill persons) can get sick.

One of the most important questions when analyzing an epidemic is the significance of early intervention. The assumed start time of interventions above was ten days. This is a very short period and, for COVID-19, is within the estimations of the incubation period. Hence, we decided to perform calculations assuming different values of delay. The results, averaged over 25 runs, are shown in Fig. 5.

The results confirm that stochastic effects are much more important for influenza-like diseases. The differences in the values of recovered persons depend very strongly on the graph creation method following the order mentioned by analyzing the epidemic curves. The introduction of groups leads initially to the worsening of results (increase in the number of infected persons) before improvement occurs (fewer ill persons when divided into 16 groups). However, the dispersion of results is so significant here than in every set of 25 runs, and a run exists in which the disease is not transferred to other individuals.

The more important conclusion comes from plot (b) of Fig. 5. Although the results for the pure BA model and those of the sample divided into 4 groups are almost indistinguishable on the logarithmic scale, these results differ by a factor of 2-3 in the same manner as in Fig. 4. The very important observation, however, is that if the assumption about much stronger links in groups is realistic, by decreasing the frequency of contact between people, we can reduce the number of infected persons by up to an order of magnitude. This effect can lead to the conclusion that greater social segregation can slow down or even stop epidemics. In this case, the stochastic effects are much stronger than for the case of a smaller number of groups, which can decrease this ratio by another two orders of magnitude.

Finally, we present the effect of taking into account the stronger links inside groups related to the real-space separation. As we wrote earlier, the percentage of individuals in groups, when divided into 16 groups, correspond to the inhabitance of 16 parts of Lodz. The plots in Fig. 6 show the percentage of people in successive areas selected from the city's total area. The outbreak starts, as usual in our calculations, in the most numerous cell. In the case considered, it is the cell in the third row and third column (see Fig. 1). One may observe that the grouping causes a larger distinction between the characteristics obtained for different areas. In the suburbs, the rate of sickness is significantly lower, except for cases where it is passed to a local hub (see lowest-right squares in plots (d) and (h)). It is essential to mention here that there exists no preference for choice, as for links in the community graph, regarding the individuals from neighboring squares. When isolated, people in less urbanized areas have an approximately ten times lower probability of getting sick.

We also emphasize an interesting effect observed when comparing some of the plots. In Fig. 6, we have four pairs of plots prepared for the same disease and intervention option but for different divisions. They are as follows: (a,c), (b,d), (e,g), and (f,h). Although the plots in pairs are usually different, one pair is visibly similar. A difference in the scale certainly exists. We can show that

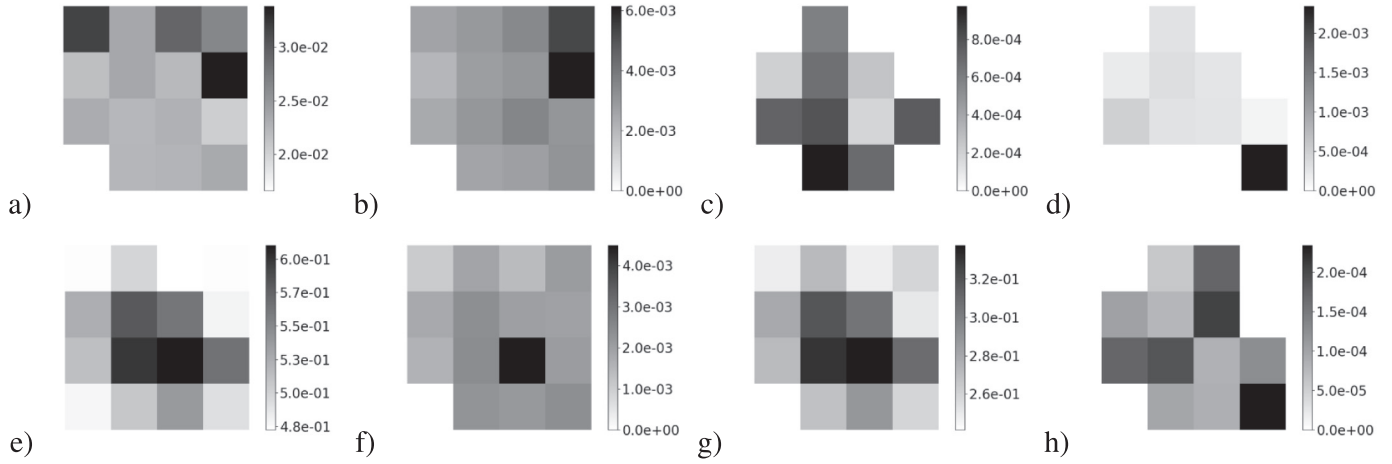

Fig. 6. The percentage of sick persons in particular areas of Lodz (see Fig. 1). Plots (a-d) present the situation for an influenza-type epidemic; plots (e-h) - for COVID-19. Particular plots correspond to the different simulation conditions: a, b, e, and f are for the pure BA model; c, d, g, and h are for the division into 16 groups (gr16); a, c, e, and g are without intervention; and b, d, f, and h are with intervention.

pair (e,g), corresponding to the COVID-19 parameters without intervention, results in the same pattern of illness in particular areas. Since the corresponding plots (f and h) with intervention are explicitly different, we can say that for this aggressive disease, an accurate description of the relations inside the community graph is more important for the correct prediction of the possible course of the disease.

4. Conclusions

Every epidemic has two components: a social one and a medical one. We presented a model that integrates both components by considering the spread of disease as the effect of individual acts. These acts correspond to the direct transfer of pathogens between two individuals in a community network, modeled with the popular social network modeling. The study shows the influence of the community structure (the structure of links concerning affiliation to particular groups). These results show the great significance of the knowledge related to a society's observations when anticipating the potential course of an epidemic.

The formulation of the model enables easy calculation of one of the basic values related to its progression - the basic reproduction number, R_0. Following the formula given in [34], i.e., $R_0 = kbD$, we can easily adapt the notions used in our paper. Because D is the duration of infectiousness, it corresponds to t_I. b is the probability of transmission by contact and equals p_I. The determination of k - the number of contacts of infectious persons per unit time - needs some explanation. It can be calculated by analyzing the average node degree of the constructed graph $< k_N >$. These calculations are not presented in this paper, but the estimation gives an interval (dependent on the size and model) from 6 to 8.4. This allows us to calculate $k =< k_N > *p_M$. Finally, the values of R_0 can be estimated as being between 0.6 and 0.84 for influenza and between 2.1 and 2.94 for COVID-19.

The above value corresponds very well to the summary given for COVID-19 in Boldog's paper [27]. We can also easily calculate the effective reproductive number (R). Usually, it is obtained by the multiplication of R_0 by the number of individuals in the S state. Here, we can add the multiplication by the p_M parameter; as a result, we can observe the effect of the intervention on R and estimate the time at which R passes the critical value $R = 1$.

In our opinion, the very important property of the proposed model is its natural way of including the stochastic character of the disease-spreading process. Applying the random approach to every possible contact, we can estimate not only the average profiles of epidemic curves but also the range in which the real number of ill persons can lie. We can try to find the crucial nodes/individuals whose identification will allow us to reduce the range of an epidemic. We can also try to estimate the possible result of this identification. Importantly, by individualizing the features of agents, we can also individualize the transfer of illness between them.

The model enables the easy introduction of different forms of intervention as well as new potential outbreaks. Considering intervention, we show the results of its typical form - isolation and quarantine. There is, however, a very simple way to introduce different procedures, i.e., vaccination, but other forms, as shown in [35], can be enlisted for the model. When entering a new illness, we have to change the attribute of the selected, either intentionally or randomly, individual.

Another important piece of information that can be estimated from the presented model is the one about the anticipated epidemic duration. By trying to recognize the inflection point (time), we can determine the time at which the disease can be significantly impaired. In our calculations, these values are limited to one per one thousand total ill persons and reaches a value approximately three times greater than the time between outbreak and inflection.

References

[1] W.O. Kermack, A. McKendrick, G.T. Walker, A contribution to the mathematical theory of epidemics, Proc. R. Soc. Lond. A115 (1927) 700–721.
[2] J.L. Aron, I.B. Schwartz, Seasonality and period-doubling bifurcations in an epidemic model, J. Theor. Biol. 110 (4) (1984) 665–679.
[3] G. Nakamura, A. Martinez, Hamiltonian dynamics of the sis epidemic model with stochastic fluctuations, Sci. Rep. 9 (1) (2019).
[4] G. Sirakoulis, I. Karafyllidis, A. Thanailakis, A cellular automaton model for the effects of population movement and vaccination on epidemic propagation, Ecol. Modell. 133 (3) (2000) 209–223.
[5] N. Ferguson, D. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, D. Burke, Strategies for containing an emerging influenza pandemic in southeast asia, Nature 437 (7056) (2005) 209–214.
[6] N. Ferguson, D. Cummings, C. Fraser, J. Cajka, P. Cooley, D. Burke, Strategies for mitigating an influenza pandemic, Nature 442 (7101) (2006) 448–452, doi:10.1038/nature04795.
[7] M. Halloran, N. Ferguson, S. Eubank, I. Longini Jr., D. Cummings, B. Lewis, S. Xu, C. Fraser, A. Vullikanti, T. Germann, D. Wagener, R. Beckman, K. Kadau, C. Barrett, C. Macken, D. Burke, P. Cooley, Modeling targeted layered containment of an influenza pandemic in the united states, Proc. Natl. Acad. Sci. USA 105 (12) (2008) 4639–4644, doi:10.1073/pnas.0706849105.
[8] N. Ferguson, et al., Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand, Technical Report, Imperial College London, 2020.
[9] S. Hoya White, A. Martin del Rey, R.S. G., Modeling epidemics using cellular automata, Appl Math Comput 186 (1) (2007) 193–202.

[10] A. Holko, M. Mędrek, Z. Pastuszak, K. Phusavat, Epidemiological modeling with a population density map-based cellular automata simulation system, Expert Syst. Appl. 48 (2016) 1–8.
[11] J. Orzechowska, D. Fordon, T.M. Gwizdałła, Size effect in cellular automata based disease spreading model, in: ACRI 2018, LNCS, 11115, Springer International Publishing, 2018, pp. 146–153.
[12] A.-L. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512, doi:10.1126/science.286.5439.509.
[13] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. 74 (2002) 47–97.
[14] P. Erdös, A. Rényi, On random graphs i, Publ. Math. Debrec. 6 (1959) 290.
[15] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.L. Barabasi, The large-scale organization of metabolic networks, Nature 407 (6804) (2000) 651–654.
[16] M. Arita, The metabolic world of escherichia coli is not small, Proc. Natl. Acad. Sci. U. S. A. 101 (6) (2004) 1543–1547.
[17] R.M. May, Network structure and the biology of populations, Trend. Ecol. Evolut. 21 (7) (2006) 394–399. Twenty years of TREE - part 2
[18] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: Structure and dynamics, Phys. Rep. 424 (4) (2006) 175–308.
[19] P. Trapman, On analytical approaches to epidemics on networks, Theor. Popul. Biol. 71 (2) (2007) 160–173.
[20] S. Zhong, Q. Huang, D. Song, Simulation of the spread of infectious diseases in a geographical environment, Sci. China, Ser. D 52 (4) (2009) 550–561.
[21] A. Ramos, P. Schimit, Disease spreading on populations structured by groups, Appl. Math. Comput. 353 (2019) 265–273.
[22] D. Balcan, H. Hu, B. Goncalves, P. Bajardi, C. Poletto, J.J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, V. Colizza, A. Vespignani, Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility., BMC Med. 7 (1) (2009) 45+.
[23] J. Hellewell, S. Abbott, A. Gimma, N. Bosse, C. Jarvis, T. Russell, J. Munday, A. Kucharski, W. Edmunds, F. Sun, S. Flasche, B. Quilty, N. Davies, Y. Liu, S. Clifford, P. Klepac, M. Jit, C. Diamond, H. Gibbs, K. van Zandvoort, S. Funk, R. Eggo, C. for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Feasibility of controlling Covid-19 outbreaks by isolation of cases and contacts, Lancet Global Health 8 (4) (2020) e488–e496.
[24] H. Han, Estimate the incubation period of Coronavirus 2019 (covid-19), medRxiv (2020), doi:10.1101/2020.02.24.20027474.
[25] N.M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A.R. Akhmetzhanov, S.-m. Jung, B. Yuan, R. Kinoshita, H. Nishiura, Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: astatistical analysis of publicly available case data, J. Clin. Med. 9 (2) (2020) 538.
[26] S.A. Lauer, K.H. Grantz, Q. Bi, F.K. Jones, Q. Zheng, H. Meredith, A.S. Azman, N.G. Reich, J. Lessler, The incubation period of 2019-ncov from publicly reported confirmed cases: estimation and application, medRxiv (2020), doi:10.1101/2020.02.02.20020016.
[27] P. Boldog, T. Tekeli, Z. Vizi, A. Dénes, F.A. Bartha, G. Röst, Risk assessment of novel coronavirus covid-19 outbreaks outside china, J. Clin. Med. 9 (2020) 571.
[28] M. Shen, Z. Peng, Y. Xiao, L. Zhang, Modelling the epidemic trend of the 2019 novel coronavirus outbreak in china, bioRxiv (2020), doi:10.1101/2020.01.23.916726.
[29] The information about lodz: https://en.wikipedia.org/wiki/%c5%81%c3%b3d%c5%ba, map of districts in lodz: https://bip.uml.lodz.pl/files/public/_processed_/2/5/csm_03_mapa-lodzi-1200_579e87dc3d.png(in polish)., 2020.
[30] J. Zhao, Y. Yang, H. Huang, D. Li, D. Gu, X. Lu, Z. Zhang, L. Liu, T. Liu, Y. Liu, Y. He, B. Sun, M. Wei, G. Yang, X. Wang, L. Zhang, X. Zhou, M. Xing, P.G. Wang, Relationship between the abo blood group and the covid-19 susceptibility, medRxiv (2020), doi:10.1101/2020.03.11.20031096.
[31] P. Schimit, F. Pereira, Disease spreading in complex networks: a numerical study with Principal Component Analysis, Expert Syst. Appl. 97 (2018) 41–50.
[32] A.L. Lloyd, Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics, Theor. Popul. Biol. 60 (1) (2001) 59–71.
[33] H. Han, Estimate the incubation period of coronavirus 2019 (covid-19), medRxiv (2020).
[34] M. Lipsitch, T. Cohen, B. Cooper, J. Robins, S. Ma, L. James, G. Gopalakrishna, S. Chew, C. Tan, M. Samore, D. Fisman, M. Murray, Transmission dynamics and control of severe acute respiratory syndrome, Science 300 (5627) (2003) 1966–1970.
[35] WHO, Managing Epidemics: Key Facts About Major Deadly Diseases, World Health Organization, 2018.