Optimal covid-19 vaccination strategies with limited vaccine and delivery capabilities; a model based on social networks

Simone Santini
Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, Spain

We develop a model of infection spread that takes into account the existence of a vulnerable group as well as the variability of the social relations of individuals. We develop a compartmentalized power-law model, with power-law connections between the vulnerable and the general population, considering these connections as well as the connections among the vulnerable as parameters that we vary in our tests.

We use the model to study a number of vaccination strategies under two hypotheses: first, we assume a limited availability of vaccine but an infinite vaccination capacity, so that all the available doses can be administered in a short time (negligible with respect to the evolution of the epidemic). Then we assume a limited vaccination capacity, so that the doses are administered in a time non-negligible with respect to the evolution of the epidemic.

We develop optimal strategies for the various social parameters, where a strategy consists of (1) the fraction of vaccine that is administered to the vulnerable population and (2) the criterion that is used to administer it to the general population. In the case of a limited vaccination capacity, the fraction (1) is a function of time, and we study how to optimize it to obtain a maximal reduction in the number of fatalities.

Categories and Subject Descriptors: Applied Computing [**Health informatic**]: ; Mathematics of Computing [**Probability and statistics**]:

Additional Key Words and Phrases: covid-19; epidemics spread; social networks; random graphs; vaccination

1. INTRODUCTION

A number of projects for the development of a vaccine for the SARS-cov2 virus have been in advanced stages of development and clinical trials since mid-2020 [Shin et al. 2020; Escobar et al. 2020; Mulligan et al. 2020; Koirala et al. 2020; Lurie et al. 2020; Ahmed et al. 2020; Jeyanathan et al. 2020], and massive vaccination has begun in most industrialized countries in late 2020 or early 2021. However, off-the-lab availability of the vaccine does not, per se, guarantee widespread immunity: the vaccine will have to be produced, distributed, and administered in great quantities, and this takes time. The dire reality of world power makes it quite likely that the early production will the distributed mainly in the richest and most powerful countries, while many countries will initially receive an amount insufficient to vaccinate more than a fraction of the population. This state of affairs poses the problem of designing a vaccination strategy: who should be vaccinated first so that the limited amount of vaccine will have a maximum impact, at least while the country is waiting for more substantial amounts to be delivered?

We develop here an infection spread model suitable for studying vaccination strategies in the presence of a vulnerable group within the population, and we use

it to begin a study of vaccination schedules under various hypotheses regarding the constraints imposed by the limited availability of vaccine or the limited delivery capabilities.

The most common models in the study of epidemic diffusion are derivations of the standard SIR model [Hethcote 2000; Acemoglu et al. 2020; Shulgin et al. 1998; Satsuma et al. 2004], which makes strong hypotheses about the distribution of the population and its interactions. In all these models, if $S(t)$ and $I(t)$ are the numbers of susceptible (healthy) and infected people at time t, the probability that a healthy person enter in contact with an infected one is proportional to $S(t)I(t)$. Because of this, the model doesn't account for the different sizes of social circles that different people may have [Barbour and Mollison 1990; Britton et al. 2007; Gonzalez et al. 2008], thereby not allowing the modeling of targeted vaccination strategies.

We define a more complete model of social interaction, which takes into account the variability of social circles, and use it to simulate and evaluate various vaccination strategies. We consider a general model and ground it using the social and epidemiological data from Spain as a test-bed[1].

The model divides the population in two groups: a more vulnerable group (conventionally referred to here as the *elderly*) and a less vulnerable one (the *young*). The parameters for the two groups have been set considering the Spanish population age 20–64 for the young, and age ≥65 for the elderly. The social connections among the young have been estimated using average social circle data [Tamarit et al. 2018]. The social connections within the group of elderly, as well as between elderly and young, are parameters that we vary in our tests.

Using this model, we test various scenarios that differ in the fraction of vaccine that is administered to the elderly as well as the policy that we use to decide which young should be vaccinated first. We also consider the case in which the vaccine delivery capacity is limited and the vaccine must be administered over an extended period of time, not negligible with respect to the spread of the epidemic. In this case, we use a genetic optimization algorithm [Mitchell 1998; Santini 2020] to design an optimal vaccination schedule.

2. THE MODEL

2.1 Social structure

Social groups are customarily modeled as random graphs in which nodes represent individuals and edges represent social interactions between individuals [Barbour and Mollison 1990; Newman et al. 2001; Newman 2005]. Let $G = (V, E)$ be such a graph, with $V = \{1, \ldots, N\}$ and $E \subseteq V \times V$. We assume that the graph is undirected without self-loops ($(u, v) \in E \Leftrightarrow (v, u) \in E$ and $(u, u) \notin E$ for all u, v), and let $m = |E|/2$ be the number of edges[2] . The *neighborhood* of a node $u \in V$ is

$$N(u) = \{v | (u, v) \in E\} \tag{1}$$

[1]The epidemiological data are derived from the daily reports from the *Ministerio de Sanidad*, published in the web page www.mscbs.gob.es/en/profesionales/saludPublica/ccayes/alertasActual/nCov/home.htm

[2]The factor 2 is due to the fact that each arc between u and v is represented twice in E, once as (u, v) and once as (v, u).

and its *degree* is $d(u) = |N(u)|$. We indicate with $\bar{d} = \sum_k d(k)/N$ the average degree, that is, the average number of neighbors of a node. Let $P_k = \mathbb{P}_u\big[d(u) = k\big]$ the probability that a random node, selected with uniform probability, had degree k. One crucial observation in most social networks is the existence of relatively few people with many contacts, and a "long tail" of many people with less and less social contacts [Barabási and Albert 1999; Wang et al. 2011; Lazer et al. 2009; Palla et al. 2007]. Specifically, many a social network exhibit a *power law* or *scale-free* distribution of the type $P_k \sim k^{-\gamma}$ where $\gamma > 1$ is an exponent that determines how "fat" the log tail is, and that depends on the specific social network [Muchnik et al. 2013; Arenas et al. 2004].

Several random graph models have been defined to mimic this behavior. The one that we use here was described in [Barabási and Albert 1999; Leskovec et al. 2008]. The algorithm creates the graph by adding one node at the time, and is controlled by two parameters, N and q, where N is the number of nodes of the final graph and q a parameter that determines connectivity. The graph is built in N steps, indexed by a parameter $t = 1, \ldots, N$. At step t, a new node is created: node number t. At this point the graph already has $t-1$ nodes. Let $d(t,i)$, $i = 1, \ldots, t-1$ be the number of neighbors of node i at step t. From the newly created node t, q edges are drawn using a *preferential attachment* (or "rich gets richer") strategy: node t is connected to node i with probability

$$P\Big[t \leftrightarrow i\Big] = \frac{d(t,i)}{\sum_{k=1}^{t-1} d(t,k)} \tag{2}$$

The algorithm that generates the graph is shown in Figure 1. The function prefRnd(V, L) generates a random element of the set V using the list $L = [d(1), \ldots, d(t-1)]$ to determine the probabilities of generating elements of V as in (2). It can be shown that this algorithm generates a power law graph with $P_k \sim k^{-\gamma}$, where the exponent γ depends on the parameter q [Leskovec et al. 2008].

```
(V, E) = random_network(n, q)
1.   V ← {1}
2.   E ← ∅
3.   d ← [0|k = 1, . . . , n]
4.   for k = 2, . . . n do
5.     V ← V ∪{k}
6.     for c = 1, . . . q do
7.       u ← prefRnd(d[1], . . . , d[k − 1])
8.       E ← E ∪{(k, u)}
9.       d[u] ← d[u] + 1
10.    od
11.    d[k] ← m
12.  od
13.  return (V,E)
```

Fig. 1. The algorithm that generates the preferential attachment graph. The value $d[k]$ keeps track of the number of neighbors of node k at each stage of the construction. The function prefRnd(V, L) picks a random element of V with the probabilities given by the elements of L via (2).

```
(V, E) = join(G_y = (V_y, E_Y), G_e = (V_e, E_e), d̄_ey)
1.   V ← V_y ∪ V_e
2.   E ← E_y ∪ E_e
3.   N ← |V_e| · d̄_ey
4.   d ← [1|n ∈ V_y]
5.   for k ← 1, ..., N do
6.     u ← unif(V_y)
7.     v ← prefRnd(V_y, d)
8.     d[v] ← d[v] + 1
9.     E ← E ∪ {(u, v), (v, u)}
10.  od
11.  return(V, E)
```

Fig. 2. Algorithm that joins two graphs through preferential attachment. It assumes that $V_y \cap V_e = \emptyset$, that is, the identifiers of the nodes for the elderly and those for the young are different, Both V_y and V_e are ordered list, allowing to maintain the correspondence between the nodes and the values in d. The function unif(V) returns and element of V chosen randomly with uniform probability.

The complete model is the union of two such graphs, independently generated on two disjoint sets of nodes (viz., the *young* and the *elderly*) and then interconnected through preferential attachment. First, two power-law graphs are generated. The first is a graph $G_y = (V_y, E_y)$ of young people generated with parameters N_y and q_y ($|V_y| = N_y$), resulting in a graph with exponent γ_y and $\bar{d}_y$ average neighbors per node. The second graph, of elderly people, $G_e = (V_e, E_e)$ is generated with parameters N_e and q_e ($|V_e| = N_e$), resulting in a graph with exponent γ_e and $\bar{d}_e$ average neighbors per node.

Finally, the two graphs are merged to form s single graph

$$G = (V_e \cup V_y, E_e \cup E_y \cup E_c) \tag{3}$$

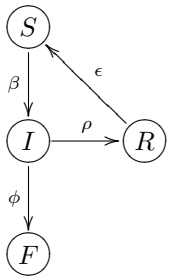
where E_c are the edges the connect the elderly and the young. To generate E_c, a number $\bar{d}_{ey}$ is selected, which will be the average number of young relations that an elderly person has. Then $N_e \cdot \bar{d}_{ey}$ edges are generated between the two graphs, selecting an elderly person at random with uniform probability, a young person using preferential attachment, and connecting them. The algorithm that joins the two graphs is shown in Figure 2. The rationale for using preferential attachment here is that certain young people (e.g., social workers, cleaning personnel, delivery people, etc.) have contact with many elders, while the majority of young people have little contact with them. The procedure allows us to create a model in which the interrelations within the groups and the relations between groups can be independently controlled. The final model is a single graph in which the nodes are partitioned in two interconnected subsets with different statistical characteristics. This is different from multi-network models [Wang et al. 2019; Boccaletti et al. 2014] in which several edge structures, with different statistical characteristics, are imposed on the same set of nodes. These models are generally used to study parallel processes with different diffusion characteristics (such as the diffusion of awareness in [Wang et al. 2019]) or the diffusion of two viruses that use different propagation vectors [Sahneh and Scoglio 2014].

$\texttt{state}_{t-1}[u]$	$\texttt{state}_t[u]$	with probability
S	I	$1-(1-\beta)^{N_I}$
I	R	ϕ
	F	ρ
	I	$1-\phi-\rho$
R	S	ϵ
	R	$1-\epsilon$
F	F	1

Table I. Probabilities of transition for a node from the states I, R, F. Transition from state S to I depends on the contact with the neighbors. N_I is the number of infected neighbors of a node.

2.2 Infection dynamics

Each node u of the graph represents an individual which, at any given time t, can be in one of four states, stored in the parameter $\text{state}_t[u]$: *susceptible* (S), *infected* (I), *recovered* (R), or *fatality* (F). Individuals transition from one state to another depending on their contacts and their recovery probability in a Markov process [Van Mieghem et al. 2008] given by the following diagram:



(4)

The only transition that depends on the interaction between nodes is $S \to I$, whereby a susceptible individual becomes infected. The probability of this transition depends on the parameter β as well as on the number of infected neighbors of a given individual. The model assumes that individuals interact with their neighbors uniformly and that they interact at all time. If a node has N infected neighbors, the probability of infection is[3] $1-(1-\beta)^N$. To implement this, given a node u in state $\text{state}_{t-1}[u] = S$ at time $t-1$, a node $v \in N(u)$ is picked randomly with uniform probability. If $\text{state}_{t-1}[v] = I$ then, with probability β the node goes to $\text{state}_t[u] = I$ and, with probability $1-\beta$, it remains in $\text{state}_t[u] = S$.

The other states evolve independently of the state of the neighbors of u, and do so according to the schema of Table I. This model exhibits, in a statistical sense, the delayed recovery and non-uniform transition characteristics of [Xia et al. 2013]. These transitions refer to non-vaccinated individuals. A separate indicator for each

[3]This is the strongest assumption that we make in out model. In reality, of course, each one of us interacts with some of our social relations more often than with others. The problem can be fixed by transforming the graph into a multi-graph, in which the number of edges between two nodes are proportional to the density of contacts between them. We currently have no good model to indicate how these multiple edges should be created. The second assumption (interaction at all time) is not so restricting: if we assume that a person interacts with only a fraction of their friends at a time, the circumstance can be modeled simply by changing the connectivity of the graph.

		d_{ey}	
		0.5	3.0
$\bar{d}_e$		α	β
1.0	A	**A**$\boldsymbol{\alpha}$	**A**$\boldsymbol{\beta}$
7.0	B	**B**$\boldsymbol{\alpha}$	**B**$\boldsymbol{\beta}$

Table II. Parameters and codes of the four different graphs we use to analyze the spread of the infection.

individual indicates whether the individual was vaccinated. Individuals can be vaccinated only if they are in state S (viz., only healthy individuals are vaccinated), and once vaccinated they will permanently stay in state S (viz., the vaccine is 100% effective–lower effectivity can be modeled by correspondingly reducing the amount of vaccine administered with respect to teh nominal amount).

3. PARAMETER SELECTION

The model depends on two classes of parameters. The *social* parameters (N_y, N_e, $\bar{d}_{ey}$; q_y and q_e, which determine $\bar{d}_y$ and $\bar{d}_e$, respectively) determine the structure of the graph, while the *epidemiological* parameters (those of (4) for young and elderly, that is, β_y, ϕ_y, ρ_y, ϵ_y, β_e, ϕ_e, ρ_e, ϵ_e) determine the spread of the infection in each neighborhood of the graph. The epidemic parameters are not observable, but they can be derived based on observable values, as shown in Section 3.2. The observables used here are relative to the early spread of covid-19 in Spain, specifically the period up to March 25, that is, the period before the lock-down decreed by the government on March 14 began to have a measurable effect effect on the spread of the epidemic (see Section 5.1 for an analysis of the criticality of these measurements).

3.1 Social parameters

The tests shown in section 4 use graphs with $N = 5,000$ nodes, with 80% of the population being young ($N_y = 4,000$) and 20% being elderly ($N_e = 1,000$). We executed preliminary tests with various values of N, observing that for $N > 1,000$ the results were stable and their qualitative characteristics did not change. The social graph of the young is fixed, generated with $q = 20$, resulting in $\gamma_y = 1.77$ and $\bar{d}_y = 16.9$. These values were chosen to be in compliance with the model in [Tamarit et al. 2018]. Then, four different graphs are generated, with different values of the social connection between the elderly ($\bar{d}_e$) and of the connection between young and elderly ($\bar{d}_{ey}$)[4]. These data result in four graphs, identified by the codes in Table II. Graphs of type **A** are characterized by sparse relations between the elderly, that is, each elderly person interacts with few other elderly people, while in graph of type **B** there are many more elderly-to-elderly interactions. Graphs of type $\boldsymbol{\alpha}$ have few cross-age interactions: each elderly person is in contact, on average, with only a few young people, while $\boldsymbol{\beta}$ graphs have denser interaction. These distinctions will allow us to analyze the effects of different vaccination strategies when used in conjunction

[4]The actual tests were carried out with 16 graphs with different combinations of these parameters. For the sake of presentation, we show only four graphs, which exemplify the different types of results that we obtained.

with other sanitary measures such as the (relative) isolation of categories at risk or the reduction of contacts between these categories and the general population.

3.2 Epidemiological Parameters

In order to determine the hidden epidemiological parameters ϕ, ρ, and ϵ from observable data, we use an SIR model as a first approximation. At the time of the simulation (December 2020), few cases of reinfection had been reported [Alizargar 2020; Kirkcaldy et al. 2020] so, as a first approximation, one can assume that the immunity is permanent in the time span considered in the simulation, and set $\epsilon_y = \epsilon_e = 0$.

Consider a population of $I(t)$ infected people maintained in isolation, so that no infected people enter the group. In the one-compartment SIR model, their evolution is described by the equations

$$\begin{aligned} \frac{dI}{dt} &= -\phi I - \rho I \\ \frac{dR}{dT} &= \rho I \\ \frac{dF}{dT} &= \phi I \end{aligned} \tag{5}$$

The solution for $I(t)$ is given by

$$I(t) = I_0 \exp\bigl(-(\phi + \rho)t\bigr) \tag{6}$$

where I_0 is the initial number of infected individuals (no new infected people enter the group, so the number of infected individuals decreases as they recover or die). The average time a person stays infected before they either die or recover is $1/(\phi+\rho)$. Assuming the day as the time unit, and an average D days duration of the disease, independently of age, one obtains:

$$\phi_y + \rho_y = \phi_e + \rho_e = \frac{1}{D} \tag{7}$$

The values of the individual parameters are determined using the measured letality, that is, the number of fatalities divided by the number of cases. Let L_y and L_e be the letality for young and old people, respectively. From (6) and the last of (5) we have

$$F(t) = \phi \int_0^t I(t)\,dt = \frac{\phi}{\phi + \rho} I_0 \Bigl[1 - \exp\bigl(-(\phi + \rho)t\bigr)\Bigr] \tag{8}$$

For large t,

$$F(t) \approx \frac{\phi}{\phi + \rho} I_0 \tag{9}$$

Therefore the letality in terms of the ϕs and the ρs can be expressed as

$$\begin{aligned} \frac{\phi_y}{\phi_y + \rho_y} &= L_y \\ \frac{\phi_e}{\phi_e + \rho_e} &= L_e \end{aligned} \tag{10}$$

Observable		Model	
Parameter	Value	Parameter	Value
D	18 days	β_y	0.1333
$R_{0,y}$	2.4	β_e	0.2667
$R_{0,e}$	4.8	ϕ_y	0.0011
L_y	0.02	ϕ_e	0.0111
L_e	0.2	ρ_y	0.0544
		ρ_e	0.044

Table III. Parameter values used in our simulation. The observable parameters in the first part are derived from the epidemic data for Spain in the period prior to March 30, before the lock-down showed its effects. The non-observable parameters in the second part are derived from those, and are those used in the model.

These equations, together with (5) yield

$$\begin{aligned} \phi_y &= \frac{L_y}{D} & \rho_y &= \frac{1}{D}(1-L_y) \\ \phi_e &= \frac{L_e}{D} & \rho_e &= \frac{1}{D}(1-L_e) \end{aligned} \tag{11}$$

The value of β is derived from its relation to the unitary infection factor R_0, which is given by

$$R_0 = \frac{\beta}{\phi + \rho} \tag{12}$$

resulting in

$$\begin{aligned} \beta_y &= \frac{R_{0,y}}{D} \\ \beta_e &= \frac{R_{0,e}}{D} \end{aligned} \tag{13}$$

The values used in the simulations of Section 4 are given in Table III.

4. METHODS

4.1 One-shot vaccinations

In the first set of simulations, we assume that all vaccination is done on a healthy population before the onset of the epidemic. In the case of an ongoing epidemic such as the one of covid-19, this corresponds to a scenario in which the still healthy population is kept effectively isolated from the people who are already infected and is vaccinated in a time very short with respect to the dynamics of the epidemic spread. We simulate the different vaccination scenarios given by the parameters specified below and then, once the stipulated number of people has been vaccinated, we simulate the unfolding of the epidemic, keeping track of the fatalities as a function of time. The simulations are characterized by three parameters:

V:. Total amount of available vaccine during the period of the simulation, expressed as a fraction of the healthy population that can be vaccinated. We are especially interested in modeling scenarios of scarcity of vaccine. Consequently, the values used in the simulations are $V = 0.05, 0.1, 0.2$.

***ψ*:.** The fraction of the available doses (viz., the fraction of V) that is administered to the elderly. The simulations use the values $\psi = 0, 0.25, 0.5, 0.75, 1$, where $\psi = 0$ means that only the young are vaccinated, while $\psi = 1$ means that only the elderly are vaccinated. The parameters are chosen so that $V \cdot \psi$ is in any case no greater than the fraction of elderly people in the population. If $V \cdot \psi$ were greater than this value then, once all the elderly have been vaccinated, the notion of "vaccination strategy" would cease to be meaningful: all the vaccine that is left after vaccinating all the elderly would be administered to the young.

***mode*:.** The criterion with which the vaccine is administered to the young. The fraction of the vaccine allotted to the elderly is always distributed uniformly inside the group. The fraction that is used to vaccinate the young is allocated according to one of the following modes:

***popular*:.** the vaccine is administered in order of "popularity", that is, the nodes in the graphs with the most neighbors are vaccinated first;

***connected*:.** the vaccine is administered in order of connections with the elderly: the young people with the most connections to the elderly are vaccinated first.

The rationale for the popular mode is that by vaccinating the people with the most connections will slow down the spread of the infection by removing dangerous "hubs". The rationale for the connected mode is to protect the vulnerable population by vaccinating the people that can spread the infection from the highly socialized young people (where the spread is rapid) to the less socialized but vulnerable elderly.

The duration of the simulations is fixed and is set to $T = 100$ days.

4.2 Incremental vaccination

The second series of simulations assumes a limited capacity to administer the vaccine, so that at most p people can be vaccinated on any given day. The total amount of vaccine available in the given period is still assumed to be enough to vaccinate a fraction V of the population (here $V = 0.2$), so the daily vaccination capacity is $p = V/T$ (expressed as the fraction of the population that is vaccinated each day).

In this case the controlled variable is a list $\psi[t]$, $t = 1, \ldots, T$, which determines, for each day, the fraction of daily vaccination capacity that is administered to the elderly. That is, on day t, $p \cdot \psi[t]$ elderly people and $p \cdot (1 - \psi[t])$ young people are vaccinated (both these values are expressed as a fraction of the total population). As in the previous case, the elderly to be vaccinated are chosen at random with uniform probability, while the young are chosen according to the selected mode.

For $T = 100$, ψ is a list of 100 values, so it clearly unfeasible and uninformative to sample it and present all the possible sequences. Instead, we use an optimization algorithm to determine the sequence $\psi[t]$ that minimizes the number of fatalities during the vaccination period. Due to the characteristics of the function to be optimized, we used a genetic algorithm [Mitchell 1998], whose details are modeled after the algorithm in [Santini 2020].

5. RESULTS

5.1 Parameter Sensitivity

The parameters of the model are derived from the epidemiological data through equations (11) and (13). The epidemiological data are subject to uncertainty, which induces an uncertainty on the parameters and, consequently, on the outcome of the model. The sensitivity of the model parameters with respect to the epidemiological data is related to the derivatives

$$\begin{aligned}
&\frac{\partial \phi_k}{\partial L_k} = \frac{1}{D} \qquad \frac{\partial \phi_k}{\partial D} = -\frac{L_k}{D^2} \quad k \in \{y, e\} \\
&\frac{\partial \rho_k}{\partial L_k} = -\frac{1}{D} \quad \frac{\partial \rho_k}{\partial D} = -\frac{1 - L_k}{D^2} \quad k \in \{y, e\} \\
&\frac{\partial \beta_k}{\partial R_{0,k}} = \frac{1}{D} \quad \frac{\partial \beta_k}{\partial D} = -\frac{R_{0,k}}{D^2} \quad k \in \{y, e\}
\end{aligned} \tag{14}$$

The estimation of the parameters is quite stable for diseases with relatively slow evolution ($D \gg 1$). For the case of covid-19, the value of D estimated through the ratio between incidence and prevalence [Irala Estévez et al. 2005] is $D \approx 18 - 20$, making the estimation stable.

On the other hand, a variation in the parameters β_k, ϕ_k, ρ_k changes the behavior of the model. Figure 3 shows the variation in the number of fatalities as a function of the variation of the parameters in several configuration with or without vaccination. The model is quite stable, exhibiting linear or sub-linear variation of the fatalities. The model will of course change if ρ_e/ρ_y changes from above to below 1 (viz., if the disease kills the young more than the elderly), but epidemiological consideration guarantee that $\rho_e/\rho_y > 1$.

These results suggest that the model is quite stable vis-à-vis uncertainty in the estimation of the epidemiological parameters. Drastic changes in the parameters (such as those deriving from applying the model to a different country or a different phase of the disease) will of course require calibrating the model to the new situation. In this case, the parameters should be computed based on the pertinent epidemiological data.

5.2 One-shot vaccination

In this section, results are presented in the hypothesis that the delivery capacity is infinite, the only limitation being the amount of vaccine available. The three parameters that affect the results are those in Section 4. On day 0, a number of young and elderly people are vaccinated following a prescribed policy and from that point on we track the number of fatality for $T = 100$ days. The infection begins at $t = 0$ with one randomly chosen young person infected (the infection never starts with the elderly). Figures 4 and 5 show the data in the hypothesis that the amount of vaccine available is enough to vaccinate 5% of the population. In Figure 4, the young have been vaccinated using the popular mode, while in Figure 5 the connected mode was used.

Each series has four graphs, one for each of the graph types of Table III, and the four curves in each graph are the fatalities for different values of ψ, the percentage

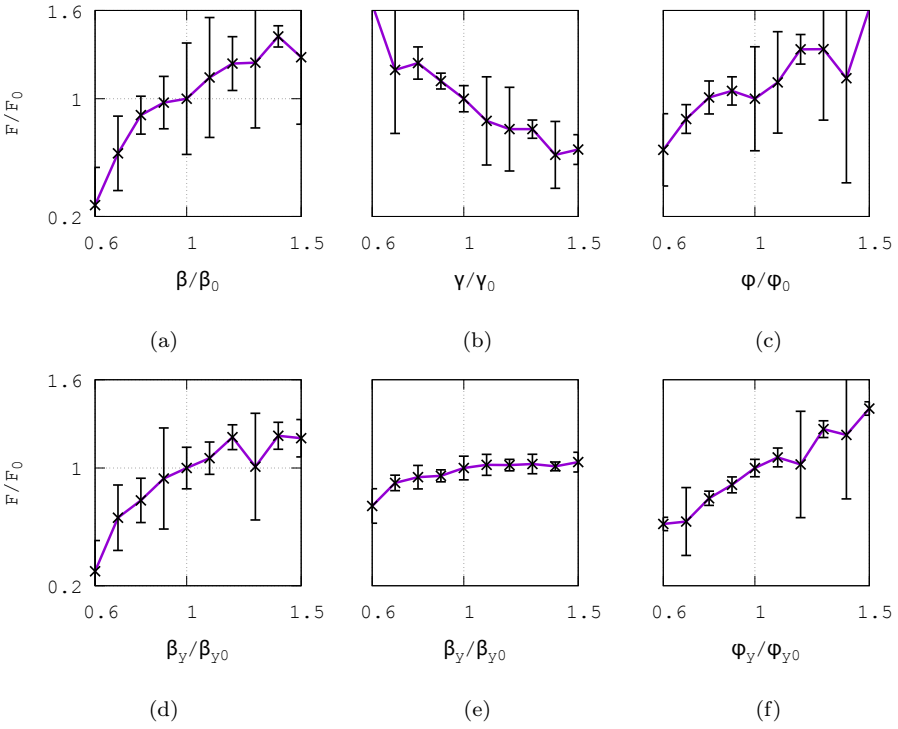


Fig. 3. Relative variation in the number of fatalities as the model parameter vary. The abscissas are relative variation of a parameter with respect to the nominal value (e.g., β/β_0, where β_0 is the nominal value of the parameter), and ordinates are the corresponding relative variation in fatalities F/F_0, where F_0 is the number of fatalities corresponding to the nominal value. The situation are the following: (a) change in β for both young and old, no vaccine; (b) change in ρ for both young and old, no vaccine; (c) change in ϕ for both young and old, no vaccine; (d) change in β for the young (no change for the elderly), no vaccine; (e) change in β for the young (no change for the elderly), 50% of the population vaccinated; (f) change in ϕ for both young and elderly, 50% of the population vaccinated.

of doses that are given to the elderly ($\psi = 0.25, 0.5, 0.75, 1$). In almost all cases, at the end of the 100 days, the result is the expected: the number of fatalities decreases as the fraction of vaccine administered to the elderly increases. The only exception is Figure 5.a, in which the differences are not statistically significant.

The smallest number of fatalities is achieved in Figure 5.a. This is the graph $A\alpha$, with little contact among the elderly and little contact between the elderly and the young. Figure 5.a is relative to the connected mode, suggesting that if the vulnerable population is relatively isolated, vaccinating the (relatively few) contacts is a good strategy.

A peculiar phenomenon, which we call *risk inversion* takes place in the popular mode, especially in the Aβ and Bβ graphs: if all doses are given to the elderly ($\psi = 1.0$) the number of fatalities in the first 20-30 days is higher than if some of the doses are given to young people. In this case, vaccinating only the elderly does

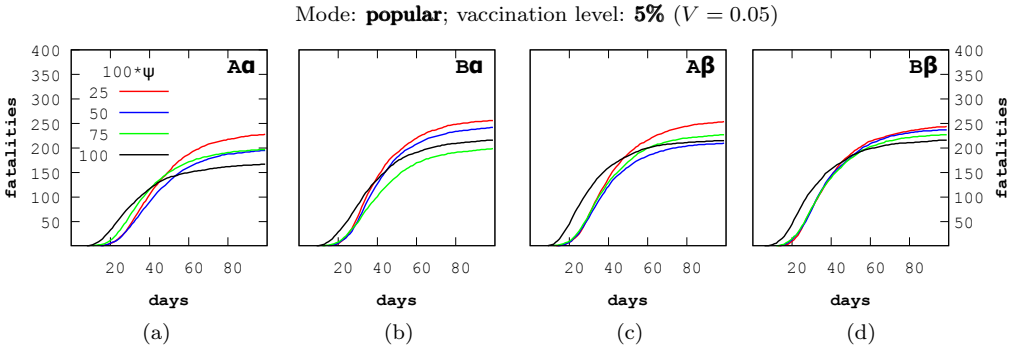


Fig. 4. Results for the infinite delivery capacity vaccination. Each curve is drawn for a given fraction of the vaccine administered to the elderly people $(0.25, 0.5, 0.75, 1)$ and tracks the number of fatalities during the 100 days of the simulation. Here, $V = 0.05$ (5% of the population are vaccinated) and the mode is **popular**.

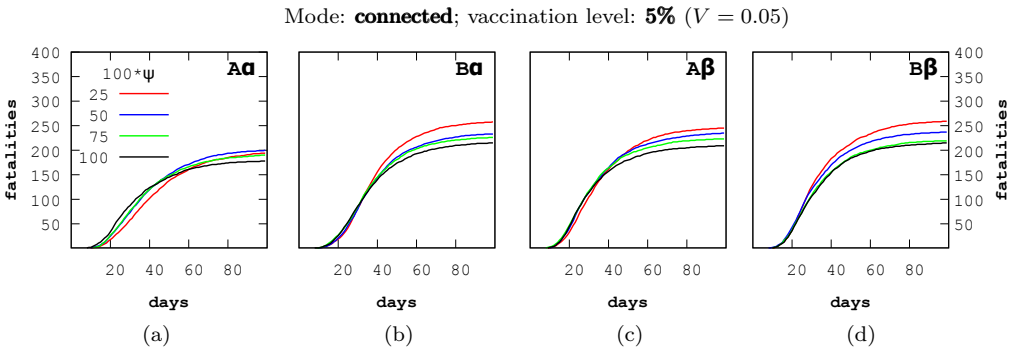


Fig. 5. Results for $V = 0.05$ (5% of the population are vaccinated) and mode **connected**. See the caption of Figure 4 for details.

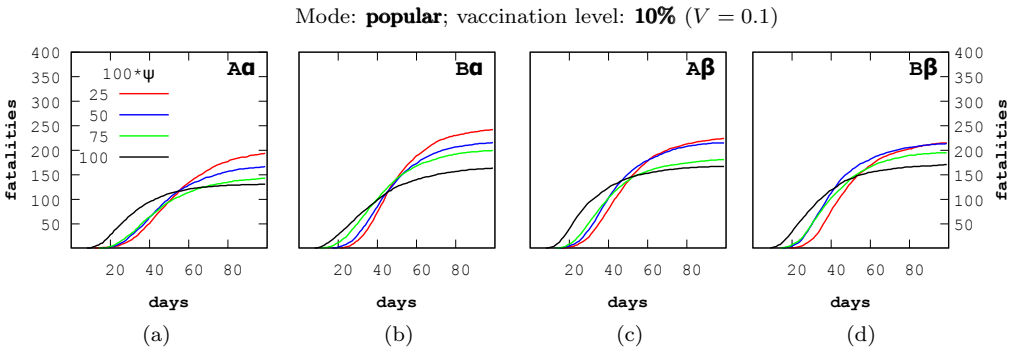


Fig. 6. Results for $V = 0.1$ (10% of the population are vaccinated) and mode **popular**. See the caption of Figure 4 for details.

offer some protection to the most vulnerable (albeit an incomplete one: there are, in the case $F = 0.05$, not enough doses to vaccinate all the elderly), but leaves the more heavily connected young completely exposed, leading to a rapid expansion of the epidemic. One the one hand this leads, in the β scenarios in which young and old are more connected, to more fatalities among the unprotected elderly and, on the other hand, to relatively many fatalities among the young. On a longer time span, the lack of protection of the elderly for low ψ compensates the effects of slow spreading, and the fatalities increase when ψ is small.

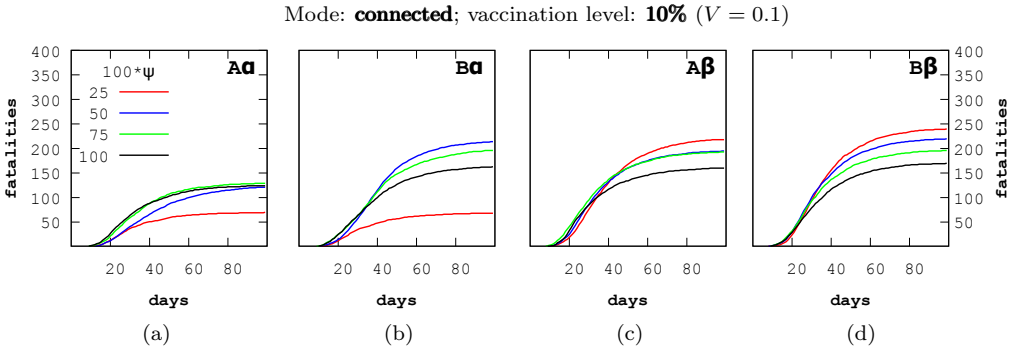


Fig. 7. Results for $V = 0.1$ (10% of the population are vaccinated) and mode **connected**. See the caption of Figure 4 for details.

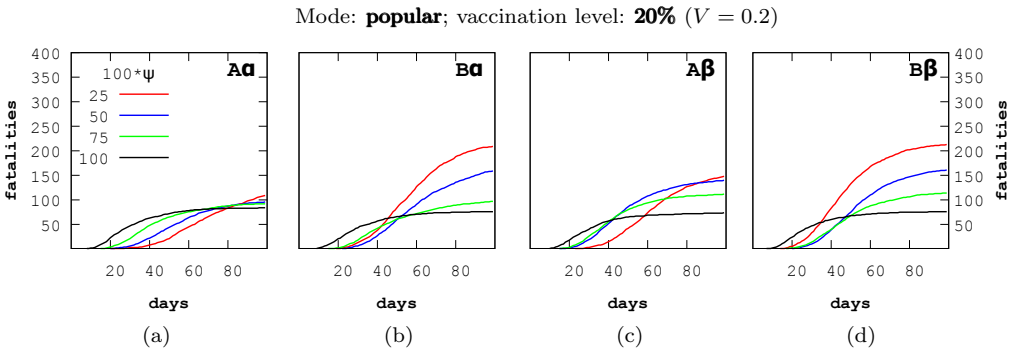


Fig. 8. Results for $V = 0.2$ (20% of the population are vaccinated) and mode **popular**. See the caption of Figure 4 for details.

Figures 6 and 7 show the results when 10% of the population is vaccinated, while Figures 8 and 9 are relative to a vaccination level of 20%. Qualitatively, most of the results are similar. It is noteworthy the sharp decrease in mortality in Figure 7.a with respect to Figure 5.a, which suggests that maintaining social distancing during the early phases of the vaccination is key to obtain the best results in a situation of scarcity of vaccine. The risk inversion is present here as in the previous cases.

One striking result is that of Figure 7.a and 7.b, in which the mortality behaves opposite as in other cases: giving only 25% of the doses to the elderly results in a

significant decrease in mortality. These figures correspond to the *connected* mode of the α graphs, with sparse connection between the elderly and the young. In this situation of high isolation and relatively high availability of doses, using 75% of the vaccine for the connected young effectively isolates the elderly while reducing the fatalities among the young. Compare this with Figures 9.a and 9.b: in this case, the number of doses is sufficient to provide a better isolation to the vulnerable group, so that in the α graphs the number of fatalities is low regardless the value of ψ. In the β graph of Figure 9.b, the stronger connection between elderly and young balances out the higher number of doses available, and the phenomenon of Figure 7.a is repeated. The convenience of targeting the "connection points" (the young people that have connections with the vulnerable population) depends on the amount of vaccine available and on the density of the relations between elderly and young.

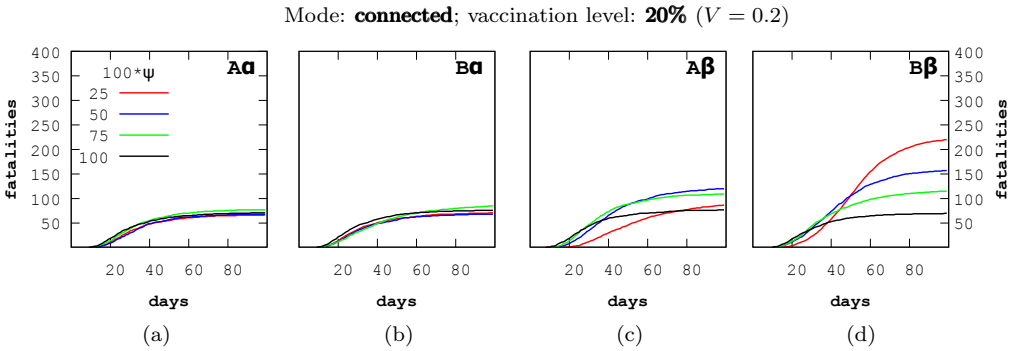


Fig. 9. Results for $V = 0.2$ (20% of the population are vaccinated) and mode **connected**. See the caption of Figure 4 for details.

The risk inversion in the early days of vaccination is present in all cases, although not with the same prominence. This phenomenon suggests that if the vaccine is available in batches, the vaccination policy might be different for different batches: a first vaccination in part to the young to limit the speed of the infection and, if a second batch is available within 20 or 30 days, a more massive vaccination to the elderly to protect them.

5.3 Incremental vaccination

Figures 10 and 11 show the optimal vaccination schedule as determined by the optimization algorithm. The vaccination schedule is assumed to extend over a period of 100 days, and the total number of doses is assumed to be sufficient to vaccinate 20% of the population. The results were obtained by a genetic algorithm using 500 generations of 800 "genes" each. The results of the best genes of several generations were checked to verify that the solutions were stable. In the figures, the black crosses mark the output of the optimizer, that is, they represent the optimal fraction of vaccine to be administered to the elderly each day. The red line is a polynomial least squares approximation of these data using a polynomial of degree 10, which allows us to visualize the general trends of the solution without

the abrupt changes: our considerations will be based on these curves. The blue and green lines are the fatalities among the young and old people respectively, expressed as a fraction of the respective population. In Figure 10 the vaccination of the young is done using the popular mode, while in Figure 11 the connected mode was used. The four graphs in each figure are relative to the four graph types of Table III. One common feature of all the strategies is the high initial fraction of vaccine administered to the elderly followed by an increase in vaccine given to the young when the number of casualties begins to grow. The peaks are more pronounced in the popular mode: in this case, after an initial short vaccination period to protect part of the vulnerable population, the focus moves to the young in order to stop the spread of infection by cutting the social hubs through vaccination. The situation is less clear in the β graphs (with many connections between the two groups). In this case, especially in the Aβ graph, the optimal strategy seems to involve a more balanced approach, with alternating focus on the elderly and the young.

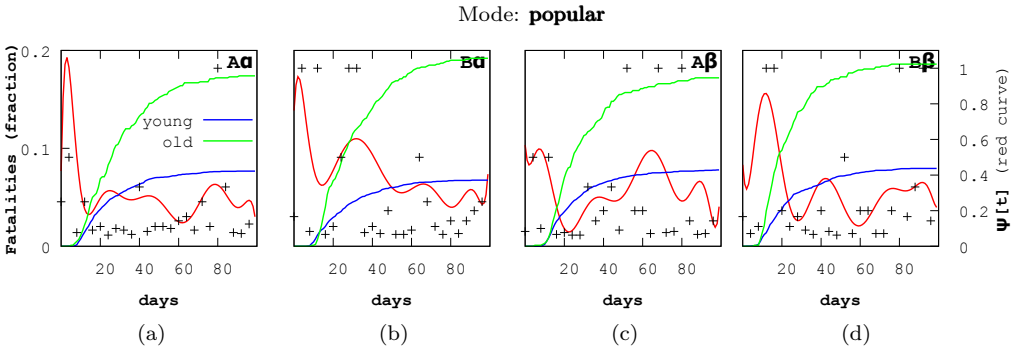


Fig. 10. Optimal vaccination strategy and the resulting mortality. The black crosses are the output of the optimizer; the red line is a polynomial approximation (degree=10) that highlights the general trends of the solution, the blue and green lines are the fraction of the young and elderly that have died, respectively, each one expressed as a fraction of the respective population. The vaccination of the young is done using the popular mode: the young people most connected are vaccinated first.

For the connected mode, in the α graphs, the peak of vaccination of the elderly is wider. Given the low value of $\bar{d}_{ey}$, relatively few vaccinations of the young are sufficient to block the spread of the infection to the vulnerable population, so the initial vaccination effort can concentrate on the vulnerable for a longer period of time. In the Aβ graph, with few elderly-to-elderly relations and strong connections with the young, the focus shifts on a much earlier vaccination of the young.

6. CONCLUSIONS

We have developed a model that, with respect to the standard SIR model, allows us to take into account the variability of the social relations of people in a social scenario. Given the existence of social groups especially vulnerable to the disease (herein referred to as "the elderly"), two factors appear crucial when defining a

vaccination strategy: the social relations among the elderly, and the connections between the elderly and the young.

The most general result that we draw from the model is that a prudent vaccination strategy should balance two requirements: protect the elderly, and vaccinate enough young people (with intense social relations) to slow down the spread of the infection.

The exact nature of this balance depends on the social connections and on who, among the young people, is vaccinated. Overall, the most effective strategy seems to be to give priority, among the young, to those who are more heavily connected to the vulnerable people (caregivers, cleaning personnel, etc.). If the amount of available vaccine is not too scarce (at least 10% of the population in our simulations), the results are good even if only 25% of the doses is administered to the vulnerable. If additional measures are taken to keep the vulnerable isolated, the 25% solution is actually the one that gives the best results. If the elderly are not isolated, then the strategy of splitting the vaccine between them and the young has positive results in a short (20-30 days) period but, in a longer period, it is necessary to move to a heavier vaccination of the vulnerable.

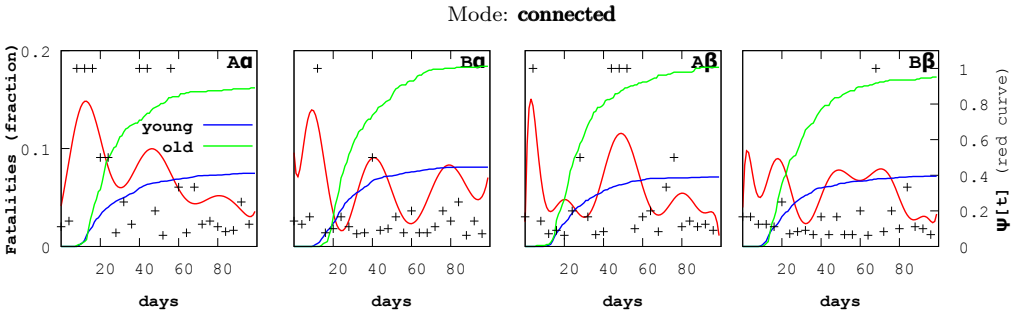


Fig. 11. Optimal vaccination strategy and the resulting mortality. The black crosses are the output of the optimizer; the red line is a polynomial approximation (degree=10) that highlights the general trends of the solution, the blue and green lines are the fraction of the young and elderly that have died, respectively, each one expressed as a fraction of the respective population. The vaccination of the young is done using the connected mode: the young people most connected to the elderly are vaccinated first.

A more detailed strategy can be devised if we take into account a limited delivery capacity and the distribution of the vaccination over the whole duration of the epidemic. In this case, the optimal strategy seems to begin with a higher fraction of elderly vaccinated then, when the number of fatalities begins to grow, with a more intense vaccination of young people to keep the spread under control. The strategy that follows this first phase is less clear, but it appear to always consist in an oscillation with periods of more intense vaccination of the elderly alternated with periods of more intense vaccination of the young, all with a general tendency to increase the fraction of vaccine administered to the young.

A few words of caution are in order. All these results were obtained through a mathematical model in which, necessarily, many of the details of the actual situation are abstracted. While the general conclusions that we derived have undoubtedly a validity, applying them in the field requires extreme caution and a deeper analysis of the evolution of the epidemic.

Also, our model has assumed that the "most popular" and "most connected" people can be identified and summoned for vaccination. This hypothesis might be reasonably realistic for the connected mode, in which the focus can often be identified by profession than with the more elusive popular mode. The distribution of the number of neighbors of the node one arrives to when crossing a random edge [Newman 2005] may help implementing this mode.

Finally, we have ignored here all the problems of public perception, and the media backlash that may be caused by a strategy that, rational as it might be, does not follow what the people and the media perceive as "common sense". These issues are clearly not in the scope of a paper like this.

REFERENCES

Acemoglu, D., Chernozhukov, V., Werning, I., and Whinston, M. D. 2020. Optimal targeted lockdowns in a multi-group sir model. *NBER Working Paper 27102*.

Ahmed, S. F., Quadeer, A. A., and McKay, M. R. 2020. Preliminary identification of potential vaccine targets for the covid-19 coronavirus (sars-cov-2) based on sars-cov immunological studies. *Viruses 12,* 3, 254.

Alizargar, J. 2020. Risk of reactivation or reinfection of novel coronavirus (covid-19). *Journal of the Formosan Medical Association*.

Arenas, A., Danon, L., Diaz-Guilera, A., Gleiser, P. M., and Guimera, R. 2004. Community analysis in social networks. *The European Physical Journal B 38,* 2, 373–80.

Barabási, A.-L. and Albert, R. 1999. Emergence of scaling in random networks. *science 286,* 5439, 509–512.

Barbour, A. and Mollison, D. 1990. Epidemics and random graphs. In *Stochastic processes in epidemic theory*. Springer, 86–9.

Boccaletti, S., Bianconi, G., Criado, R., Del Genio, C. I., Gómez-Gardenes, J., Romance, M., Sendina-Nadal, I., Wang, Z., and Zanin, M. 2014. The structure and dynamics of multilayer networks. *Physics reports 544,* 1, 1–122.

Britton, T., Janson, S., and Martin-Löf, A. 2007. Graphs with specified degree distributions, simple epidemics, and local vaccination strategies. *Advances in Applied Probability 39,* 4, 922–48.

Escobar, L. E., Molina-Cruz, A., and Barillas-Mury, C. 2020. Bcg vaccine protection from severe coronavirus disease 2019 (covid-19). *Proceedings of the National Academy of Sciences 117,* 30, 17720–6.

Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. 2008. Understanding individual human mobility patterns. *Nature 453,* 7196, 779–782.

Hethcote, H. W. 2000. The mathematics of infectious diseases. *SIAM review 42,* 4, 599–653.

Irala Estévez, J. d., Martínez-González, M. A., and Seguí-Gómez, M. 2005. *Epidemiología Aplicada*. Ariel Ciencias Médica.

Jeyanathan, M., Afkhami, S., Smaill, F., Miller, M. S., Lichty, B. D., and Xing, Z. 2020. Immunological considerations for covid-19 vaccine strategies. *Nature Reviews Immunology*, 1–18.

Kirkcaldy, R. D., King, B. A., and Brooks, J. T. 2020. Covid-19 and postinfection immunity: Limited evidence, many remaining questions. *Jama 323,* 22, 2245–6.

Koirala, A., Joo, Y. J., Khatami, A., Chiu, C., and Britton, P. N. 2020. Vaccines for covid-19: The current state of play. *Paediatric respiratory reviews 35*, 43–9.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al. 2009. Social science. computational social science. *Science (New York, NY) 323,* 5915, 721–3.

Leskovec, J., Backstrom, L., Kumar, R., and Tomkins, A. 2008. Microscopic evolution of social networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 462–70.

Lurie, N., Saville, M., Hatchett, R., and Halton, J. 2020. Developing covid-19 vaccines at pandemic speed. *New England Journal of Medicine 382,* 21, 1969–73.

Mitchell, M. 1998. *An introduction to genetic algorithms*. Cambridge, MA:MIT press.

Muchnik, L., Pei, S., Parra, L. C., Reis, S. D., Andrade Jr, J. S., Havlin, S., and Makse, H. A. 2013. Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Scientific reports 3,* 1, 1–8.

Mulligan, M. J., Lyke, K. E., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Neuzil, K., Raabe, V., Bailey, R., Swanson, K. A., et al. 2020. Phase 1/2 study of covid-19 rna vaccine bnt162b1 in adults. *Nature*, 1–5.

Newman, M. 2005. Threshold effects for two pathogens spreading on a network. *Physical Review Letters 95,* 108701.

Newman, M. E., Strogatz, S. H., and Watts, D. J. 2001. Random graphs with arbitrary degree distributions and their applications. *Physical review E 64,* 2.

Palla, G., Barabási, A.-L., and Vicsek, T. 2007. Quantifying social group evolution. *Nature 446,* 7136, 664–7.

Sahneh, F. D. and Scoglio, C. 2014. Competitive epidemic spreading over arbitrary multilayer networks. *Physical Review E 89,* 062817.

Santini, S. 2020. A random walk on area restricted search. *ArXiv 2006.14318*.

Satsuma, J., Willox, R., Ramani, A., Grammaticos, B., and Carstea, A. 2004. Extending the sir epidemic model. *Physica A: Statistical Mechanics and its Applications 336,* 3-4, 369–75.

Shin, M. D., Shukla, S., Chung, Y. H., Beiss, V., Chan, S. K., Ortega-Rivera, O. A., Wirth, D. M., Chen, A., Sack, M., Pokorski, J. K., et al. 2020. Covid-19 vaccine development and a potential nanomaterial path forward. *Nature Nanotechnology 15,* 8, 646–55.

Shulgin, B., Stone, L., and Agur, Z. 1998. Pulse vaccination strategy in the sir epidemic model. *Bulletin of mathematical biology 60,* 6, 1123–48.

Tamarit, I., Cuesta, J. A., Dunbar, R. I., and Sánchez, A. 2018. Cognitive resource allocation determines the organization of personal networks. *Proceedings of the National Academy of Sciences 115,* 33, 8316–21.

Van Mieghem, P., Omic, J., and Kooij, R. 2008. Virus spread in networks. *IEEE/ACM Transactions On Networking 17,* 1, 1–14.

Wang, D., Pedreschi, D., Song, C., Giannotti, F., and Barabasi, A.-L. 2011. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1100–8.

Wang, Z., Guo, Q., Sun, S., and Xia, C. 2019. The impact of awareness diffusion on sir-like epidemics in multiplex networks. *Applied Mathematics and Computation 349*, 134–47.

Xia, C.-y., Wang, Z., Sanz, J., Meloni, S., and Moreno, Y. 2013. Effects of delayed recovery and nonuniform transmission on the spreading of diseases in complex networks. *Physica A: Statistical Mechanics and its Applications 392,* 7, 1577–85.

A. OPTIMIZATION OF THE VACCINATION SCHEDULE

Mathematically, our problem is the following. We have a multidimensional variable, the values $\psi[t] \in [0, 1]$, $t = 1, \ldots, T$ ($T = 100$ in our case), and a function $G(\psi)$ that we want to minimize. In this case, $G(\psi)$ is the number of fatalities for the infection when ψ is used as a vaccination schedule (fraction of the vaccination that is given to the elderly each day). The value of G is not a trivial function of ψ. In particular, we cannot assume any of the standard properties (continuity, derivability, analiticity,

etc.) that most optimization algorithms assume. Because of the nature of the function, a genetic algorithm [Mitchell 1998] is a reasonable option to solve this optimization problem.

We transform the problem into a discrete one by restricting $\psi[t]$ to values of the type

$$\psi[t] = \frac{1}{n[t]+1} \tag{15}$$

with $n[t] \in [0, 15]$. Each $n[t]$ can be represented as a four-bit number, so the sequence $n[1], \ldots, n[T]$ that determines the solution can be represented as an integer of $4 \cdot T$ bits. This is the "gene" of our algorithm.

A generation is composed of a collection of G genes γ_i, $i = 1, \ldots, G$. Given a gene, we derive from it the vaccination schedule using (15) and execute the simulation. This will result in a score s_i. Standard genetic algorithms are function maximizers, so we consider as score the inverse of the number of fatalities at time T ($s_i = -V_{\gamma_i}(T)$). There are several methods to create the following generation

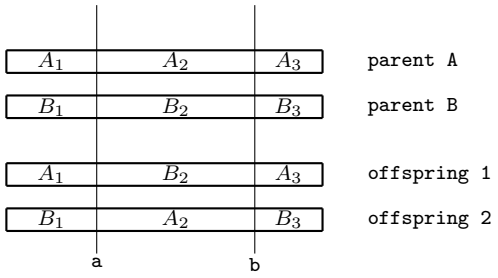


Fig. 12. Double cut for the generation of offspring. Two parents generate two offspring, mixing the bits of their genes as represented.

of individuals. Since the performance of the algorithms seems to have little dependence in the specific method used, we use one of the simplest, based on the creation of an intermediate *gene pool*. The gene pool is a set $\mathcal{P}$ of P individuals possibly replicated (generally $|P| = G$: the pool has the same size as the generations) such that the number of "copies" of an individual in the pool is proportional to its score. An easy algorithm for generating a pool is the *tournament*: we do P comparisons of pairs of individuals taken at random from the generation: the individual with the highest score goes into the pool:

```
P ← ∅
for k=1 to P do
  i ← rnd(1,G)
  j ← rnd(1,G)
  if s_i ≥ s_j then
    P ← P ∪ {γ_i}
  else
    P ← P ∪ {γ_j}
  fi
od
```

In order to build the next generation, pairs of genes are taken at random from the pool (with uniform distribution) and crossed to create two new individuals that will go into the next generation (this requires that G be even). We use the method of the *double cut* to cross the genes. Two values $a, b \in [0, 4T]$ are chosen randomly. The two offspring are then generated as in Figure 12, in which we assume $a < b$. We also define a small mutation probability: for each new gene, with a (small) probability p, we pick a random bit and flip it. Note that this method doesn't guarantee that the best individual of a generation will pass unchanged to the next, so we actually use the crossing to create $G - 2$ individuals to which we add the two best performers of the previous generation.