OXFORD

Disease spreading modeling and analysis: a survey

Pietro Hiram Guzzi, Francesco Petrizzelli and Tommaso Mazza

Corresponding Author. Pietro Hiram Guzzi, Laboratory of Bioinformatics, Dept of Surgical and Medical Sciences, Magna Graecia University of Catanzaro, Catanzaro. Tel.: +39 0961 3694148; E-mail: hguzzi@unicz.it

Abstract

Motivation: The control of the diffusion of diseases is a critical subject of a broad research area, which involves both clinical and political aspects. It makes wide use of computational tools, such as ordinary differential equations, stochastic simulation frameworks and graph theory, and interaction data, from molecular to social granularity levels, to model the ways diseases arise and spread. The coronavirus disease 2019 (COVID-19) is a perfect testbench example to show how these models may help avoid severe lockdown by suggesting, for instance, the best strategies of vaccine prioritization. **Results:** Here, we focus on and discuss some graph-based epidemiological models and show how their use may significantly improve the disease spreading control. We offer some examples related to the recent COVID-19 pandemic and discuss how to generalize them to other diseases.

Keywords: Disease Modeling, Network Models, Graph Theory, Epidemiology

Introduction

Computational modeling involves the use of algorithms and software to describe and simulate natural systems leveraging mathematics and computer science [1]. The use of computational models enables the study of the dynamics of such systems through the observation of mathematical variables that summarize their characteristics. Simulation is a valuable tool to represent the evolution of dynamical systems.

The great benefit of simulating the temporal evolution of computational models of biological systems is primarily that of anticipating or driving wet-lab experiments. Moreover, models of diseases may help predict the onset and development of diseases or forecast drug effectiveness. Many examples of such models exist at all levels of granularity: molecular processes, cell-to-cell interactions, phenotypes and transmission of diseases among people [2]. For instance, computational models are used to predict drugs' side-effects, reducing both lethal effects and years for developing better drugs [3]. Computational models are also used to design personalized medicine approaches, simulating the effect of drugs on patients based on their status and genotype [4].

We focus on the modeling of spreading of infectious diseases on this survey with the aim to track and predict their possible outbreaks [5, 6] and tailoring the strategies of intervention [7]. The use of effective computational models is a key strategy for designing and implementing interventions at all levels with the aim to block disease spreading, thereby saving lives and reducing stress on the healthcare system. Different mathematical modeling approaches have been employed to simulate the disease course:

- day-level forecasting based on time-series data;
- agent-based modeling;
- compartmental models (i.e. the partition of people into compartments labeled with their disease status, i.e. susceptible, infected or recovered) based on ordinary differential equations.

Recently, the coronavirus disease 2019 (COVID-19) pandemic has offered an unprecedented field of application for developing novel models to predict the course of COVID-19 [8]. Such models covered a broad scope, from disease monitoring to support to the individuation of ad hoc containment measures based on tailored strategies of lockdown even in the presence of vaccines [9–12]. Moreover, it should be noted that the diffusion of epidemic diseases needs contact (direct or indirect) among people. Thus, it is fundamental to consider them into the models. For airborne diseases, such as flu, Middle East Respiratory Syndrome and the recent COVID-19, transmission is almost always due to close physical interactions among people. Similarly, for sexual-based diseases, such as AIDS, the transmission is due to a sexual interaction or blood-based infection. Finally, in

Pietro Hiram Guzzi is an Associate Professor at University of Catanzaro. His research focuses on the analysis of complex networks for integration and analysis of biological data. He authored three books and wrote more than 200 scientific peer-reviewed papers. He is IEEE Senior Member and ACM Professional Member.
Francesco Petrizzelli is a PhD Student at Sapienza University of Roma. In 2018, he joined the Bioinformatics laboratory of the IRCCS Casa Sollievo della Sofferenza.
Tommaso Mazza is PI of the Bioinformatics laboratory at IRCCS Casa Sollievo della Sofferenza (FG) and CSS-Mendel (Rome), Italy. His research interests span from computational systems biology to genomics and network biology. He is member of the Bioinformatics Italian Society (B.IT.S.).

some cases, surfaces may also be a vehicle of infection [13, 14].

Consequently, the process of disease diffusion can be easily modeled using a graph data structure, whose nodes represent individuals and the edges their contacts. Simple models do not take into account any other piece of information. Refined models integrate into nodes additional information, such as characteristics of the individuals and into the edges information related to the contacts (e.g. time of exposure, environment in which contacts take place, etc). In particular, time-related information is used to build temporal networks where the edges store the duration or the times of the contact [15, 16]. Some previous works, as e.g. [17–22], have demonstrated that the use of a graph-based model may be helpful to describe the virus spreading effectively. Such a model applies algorithms borrowed from social sciences to epidemiological graphs to measure their cohesiveness, namely the importance of nodes removal in the breach. In the epidemiological ecosystem, nodes are not actually pulled out of a graph but made ineffective as a response to immunization [23]. From a semantic point of view, *node removal* and *node that is made ineffective* are the same things. Despite this, from a topological point of view, node removal causes a global change in the network structure since it causes the removal of the incident edges. In contrast, node inactivation can be easily simulated by changing a label (or a color) on the node that does not modify the network structure.

Hence, it is then possible to investigate some ad hoc policies of containment (e.g. the use of face masks, localized lockdown or tailored vaccination strategies) by such a model. The translation of these interventions into graph properties is strictly related to the underlying graph model. For example, if a simple unordered graph is used, then the wearing of masks will be modeled as edge removal. Diversely, if a temporal graph is used, then the wearing of the mask will be modeled as a change in the property of the edge, as suggested.

In recent years, the spreading of information (or more precisely, the spreading itself), the best conditions for spreading and the emergence of super-spreaders have been largely investigated in network science considering the spreading of dangerous viruses, pathogens (and misinformation or gossip) [24–26]. In [27], authors investigated the impact of the community structure of the network on percolation, simulating the spreading of an epidemic modeled through a classical susceptible-infectious-removed (SIR) epidemic model. Authors concluded that within communities spreading is critically related to denseness [28, 29], while the inter-community edges are the most critical factor for the spreading of an epidemic, regardless of the community size and shape. In particular, in [30], a study of epidemic spreading using the adjacency matrix of a graph has been proposed. In such a work, given a contact matrix (without any constraint on the structure of the matrix),

and an epidemic modeled using a susceptible-infected-susceptible (SIS) model described by two parameters β and δ describing, respectively, the rate of novel infected and the rate of recovered, it has been demonstrated an upper bound to the epidemic given by

$$\frac{\beta}{\delta} \leq \frac{1}{\lambda_{max}}, \tag{1}$$

where λ_{max} is the largest eigenvalue of the adjacency matrix. Consequently, the property

$$deg_{avg} \leq \lambda_{max} deg_{max}, \tag{2}$$

where deg_{avg} and deg_{max} are the average and the maximum node degree, respectively, holds.

An intuitive explanation of this concept follows this consideration: the intrinsic characteristics of the epidemics (represented by β and δ) cannot be changed since they represent the characteristics of the virus. We can act on the structure of the contacts that impact the λ_{max}. Removing edges, which is realized by preventing contacts or immunizing individuals, causes the diminution of λ_{max}; thus, the right part of the equation grows, thereby controlling the epidemic.

We here present a synopsis of the main models used to analyze disease spreading. We start discussing classical compartmental models, then we present contact-based models. Finally, we present an application of a contact-based model on the prioritization of vaccination strategies for COVID-19.

Compartmental models

Among the others, compartmental models are probably the most used in the mathematical modeling of infectious diseases. The investigated population is assigned to compartments identified by labels. Such models study the evolution of the population that belongs to each compartment. Examples of such compartments are SIR, (Susceptible, Infectious or Recovered) or SIS (Susceptible, Infectious or Susceptible). An SIR model is usually described by the following system of ordinary differential equations:

$$\begin{cases} \dfrac{dS}{dt} = -\dfrac{\beta IS}{N}, \\[2mm] \dfrac{dI}{dt} = \dfrac{\beta IS}{N} - \gamma I, \\[2mm] \dfrac{dR}{dt} = \gamma I, \end{cases} \tag{3}$$

where N is the population size, and $\frac{dS}{dt}$, $\frac{dI}{dt}$ and $\frac{dR}{dt}$ represent the variation of the susceptible, infected and recovered population, respectively. The spreading of a disease and the impact of the containment measures are summarized by the β and γ parameters. Given such a model, it may be simulated through the use

of ordinary differential equations (ODE) solvers or the original Gillespie's Stochastic Simulation Algorithm [31].

Starting from this initial scheme, a set of different models have been derived. In particular, it has been observed that many common diseases such as flu or other infectious diseases do not confer durable immunity. Therefore, people that are infected become susceptible after recovering. These diseases can be modeled using an SIS model [32], which can be formalized as follows:

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta SI}{N} + \gamma I \\ \frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \end{cases} \tag{4}$$

A notable variation of the SIR model exists that includes the death of individuals as parameter (SIRD):

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta IS}{N}, \\ \frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I - \mu I, \\ \frac{dR}{dt} = \gamma I, \\ \frac{dD}{dt} = \mu I, \end{cases} \tag{5}$$

Other variations were proposed to add the possibility to model the vaccination impact, considering or not the death event (i.e. SIRDV and SIRV models) [33].

$$\begin{cases} \frac{dS}{dt} = -\frac{\beta(t)IS}{N} - \upsilon(t)S, \\ \frac{dI}{dt} = \frac{\beta(t)IS}{N} - \gamma(t)I, \\ \frac{dR}{dt} = \gamma(t)I, \\ \frac{dV}{dt} = \upsilon(t)S. \end{cases} \tag{6}$$

A synopsis of such models is depicted in Figure 1.

Such models are usually formalized and simulated through the use of ODEs or using stochastic frameworks. In this way, the characteristics of the population are described by the parameters of the equations. They are effective when used to predict aspects as how a disease spreads, the total number of infected people, the duration of an epidemic or to estimate epidemiological parameters such as the reproductive number. They have the potential to show how specific public health interventions may affect the outcome of an epidemic, e.g. what is the most efficient technique for issuing a limited number of vaccines in a given population [34–36] or evaluate the impact of non-pharmaceutical interventions on the viral spread [37, 38].

Applications of Compartmental Models

Compartmental models have been used in the past to model the spreading of flu [39]. Coburn et al. surveyed some of the main models focusing, in particular, on the future of swine flu (H1N1) [40]. Such models mainly used data about the flu illness derived from public databases or, more recently, from social network data [41, 42].

Today, the recent COVID-19 pandemic has given researchers the possibility to test existing models and define novel models fed by constant data and motivated by an extreme emergency. Giordano et al. [43] proposed a refinement of a compartmental-based model that considered eight compartments (or stages of infection): susceptible (S), infected (I), diagnosed (D), ailing (A), recognized (R), threatened (T), healed (H) and extinct (E), i.e. a SIDARTHE model. This model has been tested in different scenarios characterized by different policies of containment and availability of Intensive Care Unit (ICU) beds. It was able to predict both the trend of the infected people and the need for ICU beds. The authors also evaluated the effectiveness of implementing different lockdown measures: moderate versus strong lockdown and distancing measures. As stated by the authors, the aim of the models was to provide policymakers with a tool to assess the consequences of possible strategies. Kohler et al. [44] used the SIDARTHE model in the German context. They proposed an optimal open-loop and feedback control strategies to handle the outbreak. The model has been later extended into the SIDHARTE-V model, with the aim to model the vaccination effects [45]. It is depicted in Figure 2.

Similarly, Oliveira et al. [46] developed a compartmental model based on the following eight compartments: susceptible (S), exposed (E), infectious (I), split into asymptomatic/non-detected infections (I_a) and symptomatic (I_s), hospitalized (H), requiring ICU admission (U), recovered (R) and deceased (D). The authors used the model to study the pandemic in Bahia, Brazil, considering a population of 14.8 million individuals.

Summary

Compartmental models have been used for a long time to describe the spreading of many infectious diseases, from plague and black death outbreaks to SARS and flu [34]. More recently, they have been used to support policy decision-making around the COVID-19 epidemic.

All the compartmental models share the same underlying theoretical framework: they partition the population into different compartments such as susceptible, exposed, infectious and removed and track the movements of individuals between these groups. The transitions among groups are based on the homogeneous mixing hypothesis; therefore, each individual has the same probability of meeting others. Ordinary differential equations describe such transitions, while continuous-time Markov chains (CTMC) are also applied in some cases. Individual- or agent-based models are sometimes used to represent individuals, thus realizing fine-grained
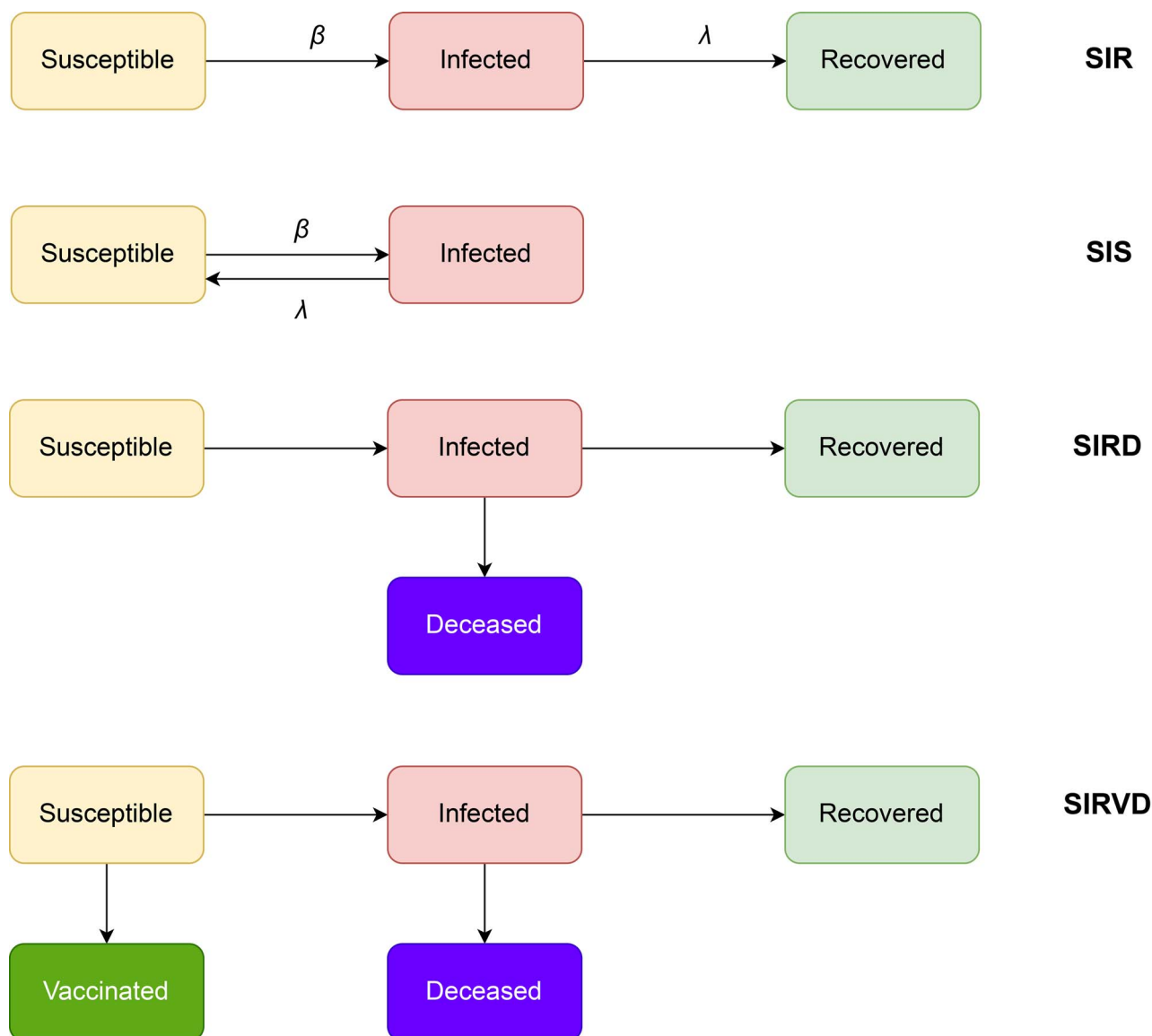
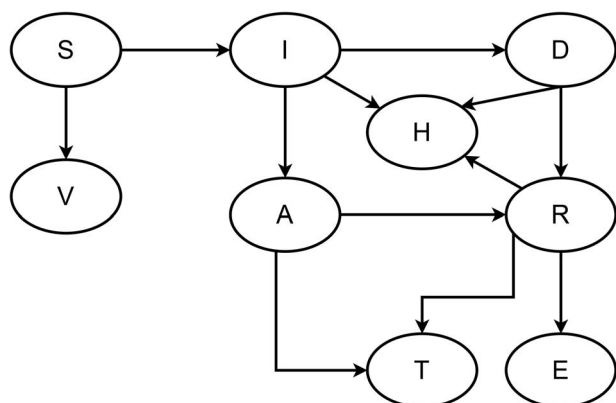**Figure 1.** A synopsis on compartment models.

Figure 2. The SIDHARTE-V Model and related state transitions. The model distinguishes among infected people on many compartments: I, infected (asymptomatic or pauci-symptomatic infected, undetected); D, diagnosed (asymptomatic infected, detected); A, ailing (symptomatic infected, undetected); R, recognized (symptomatic infected, detected); T, threatened (infected with life-threatening symptoms, detected); V, vaccinated.

modeling of the characteristics of each one, such as different contact patterns or susceptibilities to the disease. Despite the clear advantages of such models, it should be noted that they require high-performance computing infrastructures. [47–50].

Contact-Based Models

All the previous models were based on the assumption that the contact structure (also called 'mixing pattern') is homogeneous. However, this hypothesis is often not verified in real scenarios [51]. Let us consider contacts among individuals; the probability of contacts among friends, coworkers and familiars is trivially higher. Moreover, people with more social relationships have more contacts, thus they are more likely of being infected and infecting others. Historically, the homogeneous mixing assumption was partly motivated by the fact that data
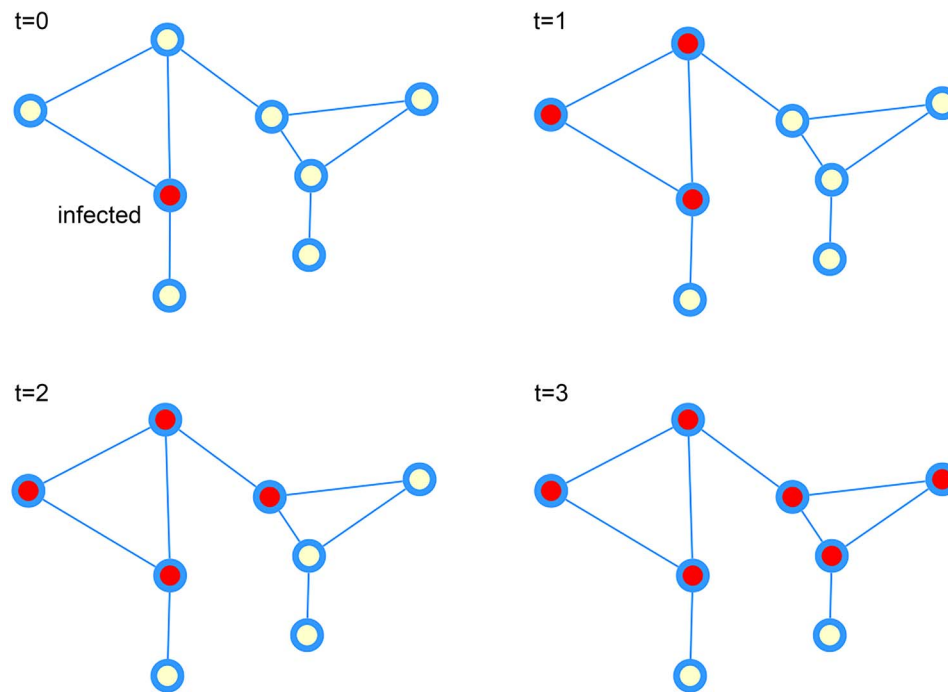
Figure 3. In contact-based models, the population is modeled as a graph, whose nodes represent individuals and edges contacts among them. In such a model, spreading is easily represented by a Markovian model or by an ODE system. It is immediate that spreading now depends both on the characteristics of the virus and on the structure of the contact network. The figure represents a toy spreading process that originated from the red node. It is worth noticing that the starting point of the spreading may generate different outcomes and that high-degree nodes are more relevant for disease spreading (and controlling).

about mobility and contact among individuals were hard to produce and store. The inefficiency of such models to correctly simulate the disease spreading has been discussed in the literature. For COVID-19, Moein et al. [52] demonstrated the relevant error in forecasting COVID-19 cases.

Recently, the availability of mobility and contact data, produced for instance by cellular phones, with a very high temporal frequency, led to the introduction of novel models gathering information from these data and offering the possibility to extend previous results on virus spreading on the Internet [53]. The main feature of these models is the integration of a diffusion model (based on one of the previous ones) to a contact pattern based on temporal networks (Figure 3). In such networks, nodes represent the individuals and edges the contacts among them [54]. Each node has a label associated corresponding to its state in the model, e.g. infected, susceptible and recovered. In such a way, contact and mixing are related to the network structure, and transitions among states are related to the transition probability of the compartmental model and the effective contacts [55, 56].

The literature contains some relevant examples of the use of such models spanning a large class of fields of application, from the epidemic of zoonosis to the viral epidemic in humans. Darbon et al. [57] used a network-based approach to analyze the outbreak of brucellosis, a highly contagious zoonosis of vaccines. In particular,

the network model represented cattle trade movements from database data and reported the estimation of the vulnerability of a cattle market to brucellosis. The use of the model showed high accuracy in predicting the number of bovine brucellosis outbreaks. Marziano et al. used retrospective data about varicella infections, integrated with demographic data and social contact patterns [58] to model the interplay between the increase of contacts of young children and varicella transmission dynamics. Chang et al. [59] first demonstrated the effectiveness of the use of mobility data for building contact networks in the US metropolitan areas. They extracted mobile phone data to map the movements of 98 million people. All these networks were integrated into a single SEIR model. Results showed that the obtained model could reproduce human behaviors and make accurate predictions on cases. A similar approach has been applied in [60] to model the impact of human mobility networks on the global spreading of the virus. Authors found that most countries had similar levels of virus spreading, and few countries acted as *super-spreaders*. More recently, Hartnett et al. [61] used mobile phone data to study the outbreak of COVID-19 in Portland (Oregon). They compared a contact-based model with a compartmental SEIR model based on a contact network. Their findings evidenced that predictions based on contact networks are more accurate and reliable. Moreover, targeted vaccination of hub nodes in the network resulted in faster epidemic control. Also, considering the Brazilian scenario, Peixoto

et al. [62] used a contact-based model in early 2020 looking at Sao Paulo and Rio de Janeiro states. The overall framework was based on geolocation data and a diffusion model. They observed a stratified risk depending on the structure of the contact networks and identified as high-risk the regions closer to capital cities. Muller et al. [63] combined data from a transportation model based on person-centric data-driven human mobility data, an SIR model and a person-centric disease progression model. A transportation model is used to reconstruct patterns of mobility of the individuals. Then, such patterns are used to model contacts and then derive the conditional probability of being infected given a contact with an infected one. Transportation models represent a way to reconstruct contacts, thus they enable the possibility to derive the network of contacts, without assuming a uniform mixing. They considered many parameters, such as the room sizes, air exchange rates, disease import, changed activity participation rates over time (coming from mobility data), masks, indoors versus outdoors leisure activities and contact tracing.

Epigraph [64] is a parallel simulator based on agents that can perform stochastic simulation. The tool was initially developed to simulate flu propagation, and recently it has mainly been used to model propagation of COVID-19 disease across vast geographic expanses. It is based on a contact model structured on an interconnection network among individuals based on actual interactions. Data related to the people interaction are extracted from existing databases of social networks and demographic data. The current version of the model leverages information from the characteristics of each individual, such as work, school, home and during leisure time, and the mined relationships. Finally, the spatial dynamics of the propagation of the virus is modeled using a transportation model. Epigraph includes climate and meteorological factors, such as temperature, atmospheric pressure and humidity levels. It also includes in its analyses some specific characteristics of the virus variants and the impact of both non-pharmacological (distancing policies, lockdown) and pharmacological interceptions (the Pfizer-BioNTech, Moderna, Astra-Zeneca and Janssen vaccines). Covasim [65] is an open-source model developed to simulate coronavirus spreading. It includes country-specific demographic information and different transmission models in heterogeneous social contexts (e.g. home, schools, workplace). It also considers phenotypic information about people (i.e. age-specific effects) and genotypic characteristics (e.g. different viral dynamics). It offers the user the possibility of inserting non-pharmaceutical interventions (such as physical distancing and protective equipment) into pharmaceutical interventions (vaccination, testing). It has been applied to simulate pandemics in many countries to support policymakers.

Kucharski et al. [66] proposed a mathematical model based on simulation and individual level transmission.

Contacts were stratified on the basis of context (household, work, school or other), and data were extracted from BBC Pandemic data of 40 162 UK participants. The model aimed to study the impact of combined isolation and tracing strategies to reduce virus spreading. Saunders et al. [25] presented a network model of the transmission of COVID-19 able to study the immunity based on different vaccination scenarios. Authors found that the average immunity length after an infection is one of the most critical parameters to control the spread of the disease. In parallel, they also found that people's choice to be vaccinated impacts the control of spreading. The vaccination of highly connected individuals first is the quickest strategy for controlling the diseases. Simulations are based on the integration of a network model and a Susceptible-Infected-Recovered-Susceptible (SIRS) model. Despite this, one limitation of this paper is that authors use only a scale-free model of contact and do not use real data.

Many other works used a compartmental model to study the disease spreading [34, 67–70] and the impact of the vaccination [71–74]. Google Scholar returned more than 48 000 items by searching for *COVID19 spreading modeling*.

Summary

Contact-based models rely on the accurate qualitative modeling of the following scenarios: from the actions that aim to investigate the number of positives and their contacts (e.g. testing of people and contact tracing) to the consequences of the results of the previous actions that impact to the structure of the model (e.g. lockdown, use of the masks, etc.). The previous consideration represents one of the main advantages of contact-based models with respect to classic compartmental models. They can simulate some interventions, but they cannot include in their analysis the contact tracing of individuals that rely on the knowledge of the network structure.

Some previous works used some approaches to integrate the structure of the contacts in a compartmental model. As already noted in [34], these models fail to capture the different impacts of different contacts and how contact tracing may help in containing outbreaks [75]. Such models assume that spreading characteristics depend on the relevant contact network's properties (e.g. degree distribution, intrinsic node properties). Consequently, they rely on the corollary that the structure of the network is not random; in that case, the predictions of network epidemics are indistinguishable from the predictions of the traditional epidemic models, as reported by A.L. Barabási in [76].

A Case Study on Modeling COVID-19 Vaccination Strategy

The computational models find a natural field of application in predicting the impact of vaccines when considering that the production of vaccines has a lower rate than

TABLE 1. Landscape of contact-based models

Model	Aim	Data
Darbon et al. [57].	Zoonosis	Not Available
Muller et al. [63]	COVID Simulation	Available at Author Website.
Hartnett et al. [61]	COVID Spreading and Vaccination	Available at https://github.com/RANDCorporation/network_vaccination
Epigraph [64]	General Purpose Simulator	Preloaded Data
Kucharski et al. [66]	Diffusion Modeling and Control	BBC Pandemic data from 40 162 UK participants
Covasim [65]	General Purpose Simulator	Pre-loaded with data as reported by the UN Population Division 2019 (population.un.org).
Saunders et al. [25]	Simulation of Vaccination Strategy	Simulated data available at https://github.com/HAHerrmann/NetworkEpidemics

infections. Therefore, there is a need for developing ad hoc prioritization strategies [77] both in Western countries and in those with low availability of vaccines (e.g. the developing countries). Furthermore, there is evidence that COVID-19 has different impacts and transmission rates by age, sex and social groups [78–80], showing a higher fatality rate in older people and men [81] and in the presence of comorbidities. Consequently, different strategies have been discussed and evaluated by the scientific communities to support the governments [82]. These strategies consider as variables the characteristics of vaccines (e.g. the possible variation of the mitigating impact caused by aging) and the effect of COVID-19 on social groups (e.g. fatality rates by age, transmission rate by worker class). The objective of the models has often been the minimization of deaths and spreading [82].

Goldstein et al. [82] demonstrated that vaccination of the oldest enabled us to save both the most lives and most years of life. Older people's prioritization was chosen as the main criterion in many countries such as Italy and the United States. These countries also gave priority to healthcare professionals, teachers and caregivers. Common sense suggests that a good prioritization scheme should choose the best trade-off between saving the maximum number of lives and the most future life. Despite the effectiveness of this approach, it has been clear that a vaccine allocation strategy requires the incorporation of a model of transmission and the epidemiological characteristics of the disease among social groups [43, 83, 84]. This has also been a counterpart in some recent works that demonstrated the inefficiency of classical epidemiological models in forecasting COVID-19 trend [52]. Buckner et al. developed a framework for the optimal dynamic prioritization of vaccines integrating the transmission dynamics and COVID-19 characteristics into their model. The authors used a compartmental model for modeling COVID-19 spreading. They studied the impact of optimizing three objectives: minimizing novel cases, years of life lost (YLL) or the number of deaths. The first contribution of this work was that no one strategy fits all benefits, i.e. the currently limited knowledge of many variables makes it impossible to

develop a system that can simultaneously achieve all three objects. Second, the authors demonstrated through three different simulations that the precedent objectives require clearly different strategies. Minimizing mortality (YLL or deaths) requires the vaccination of older essential workers (especially 60+ years). Alternatively, the minimization of the number of cases is achieved by prioritizing essential workers. Authors clearly state that policies that differentiate and target essential workers in addition to age substantially outperform those considering the age alone. In [85], Jentsch et al. discussed the problem of optimizing the strategy of vaccination. They demonstrated that a strategy based only on age is not optimal compared with a contact-based strategy. They developed a mathematical model working on data collected in Ontario and simulated the impact of different vaccination strategies. The contact strategy was based on the allocation of vaccines according to the relative role played by different age groups in the transmission. This strategy prioritized ages 15–19 years first, 20–59 years second, and gave the least priority to older or younger ages.

The analysis of these approaches suggests, to the best of our knowledge, three primary considerations:

- the optimization strategy depends on the desired goals;
- the integration of the characteristics of the modeling improves the performance;
- a dynamic strategy may outperform a static one.

Despite this, we retain that modeling the viral spreading using only a classical SEIR approach may not be the best choice since some parameters are considered globally, while the spreading involves individual contacts. Compartmental models, such as the SEIR model, are based on the homogeneous mixing hypothesis, i.e. each individual has the same probability of being connected with any other individual. This strong hypothesis leads to the introduction of a bias that is relevant for real case scenarios.

Hence, the task of sorting relevant groups of individuals, i.e. nodes that are topologically important in a graph, is of paramount importance. This aim can be assessed
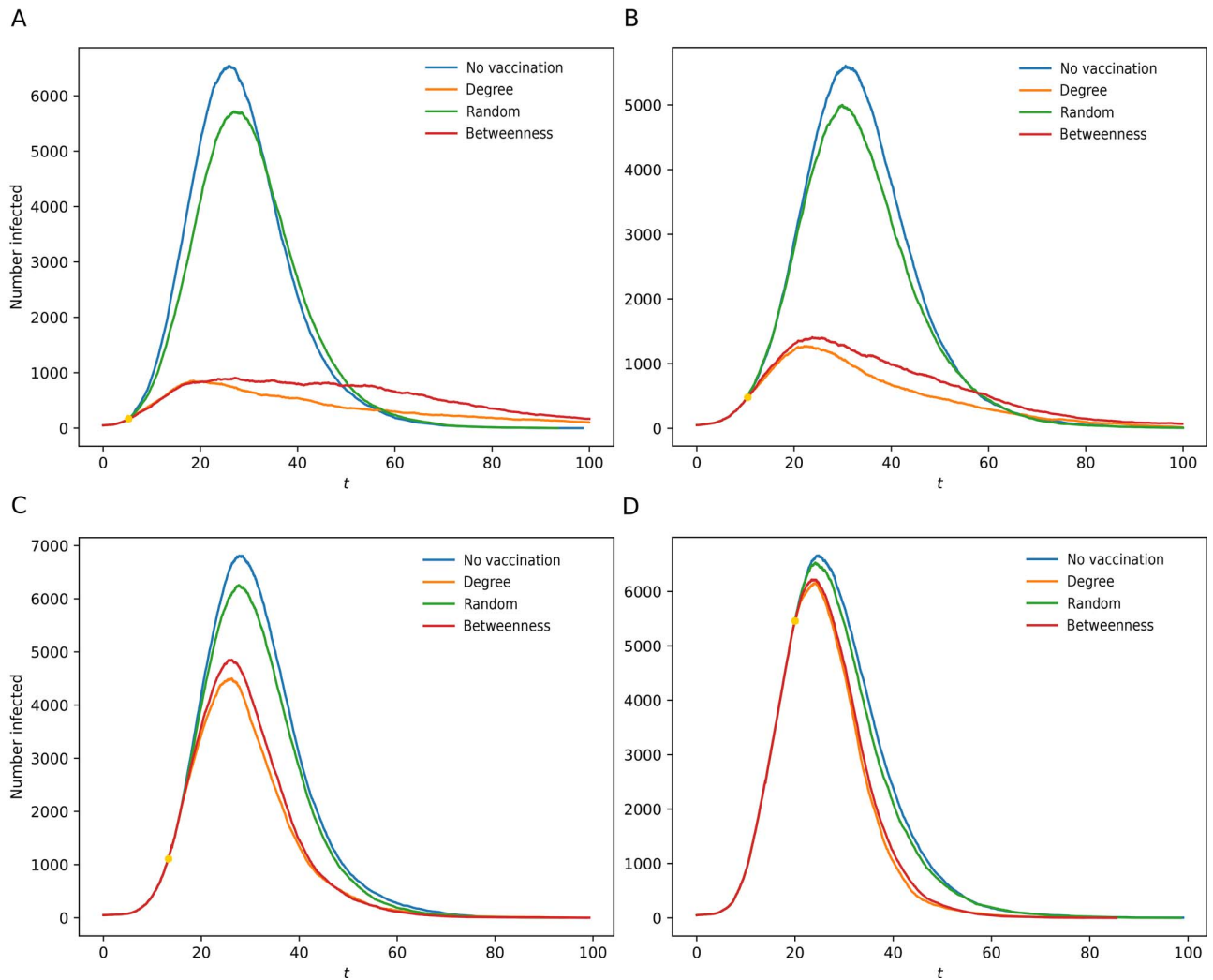
Figure 4. Example of different topology-based vaccination strategies applied at different times (A=5t, B=10t, C=15t, D=20t) in independent populations represented using a Duplication Divergence generative model. Both a degree-based vaccination (orange) and a betweenness-based vaccination (red) were compared with a random vaccination (green), removing 1000 top-degree nodes, 1000 top-betweenness nodes and 1000 random nodes, respectively. As a control, the spreading of the infection without any vaccination was simulated (in blue). The yellow dots indicate the different vaccination times.

by several topological and well-established metrics. To mention a few, we refer to *degree* and *betweenness* centrality metrics, which measure the importance of a node based on the number of their direct contacts or the extent of their involvement in social contact traces. For these two metrics, implementations to assess the centrality of (groups of) individuals exist [86, 87]. Furthermore, more advanced and focused metrics were conceived to specifically tackle the problem of identifying important groups of nodes. In particular, the KPP-NEG metrics was defined by S.P. Borgatti in [23] with the explicit aim to measure importance in the breach, namely the amount of reduction in cohesiveness of a graph that would occur if the nodes were not present. This concept can be declined in the public health context as *the need to select a subset of population members to immunize or quarantine in order to optimally contain an epidemic*. These are implemented in several software packages, e.g. Cytoscape [88] and Pyntacle [89], and easily embeddable in simulation frameworks.

As an example, we generated artificial networks with known topological configurations. In particular, we created networks of 25 000 nodes, i.e. individuals organized according to the *Duplication-divergence* generative model. Each network corresponded to an adjacency matrix, of which we calculated the largest eigenvalue before and after the removal of selected nodes (cf. Eq. 1). In fact, we deprived the networks of either k random nodes or k top-scored nodes according to *degree* or *betweenness* centrality scores. These two metrics are the most studied and effective targeted vaccination strategies: a *degree*-based vaccination strategy aims to remove individuals with a high number of social interactions, while a *betweenness* centrality-based approach breaks the network into smaller groups of individuals, thus limiting the spread of the infection from one group to another. We then calculated the λ_{max} reduction values obtained with both node elimination strategies. It is worth noticing that node removal roughly mimics the vaccination of an individual. The decrease obtained when removing the

most topologically central nodes was higher in all cases (Figure 4).

In detail, we chose a transmission rate such that it takes about 14 days for the disease to go from 50 to 500 infected, and we applied our vaccination strategy at different times (t=5,10,15,20) in independent simulation traces. In all the simulated cases, we observed a substantial impact on the transmission (i.e. the number of infected) when the top-degree and top-betweenness nodes were removed compared with the slight reduction of the number of infected using a random vaccination strategy. Indeed, these strategies are much more effective as earlier the vaccination occurs. Similar results were obtained if using other generative algorithms (i.e. Erdoős Rényi and Barabási-Albert, data not shown). The simulation framework used in this work was the SIR as implemented in the Python library *Epidemics on Networks* (EoN) [90].

Conclusion

Modeling disease spreading is a crucial research area in computational epidemiology. Recently, the COVID-19 pandemic has motivated the introduction of novel approaches based on the integration of Graph Theory and classical ordinary differential equations. These models overcome the classical hypothesis of homogeneous mixing among people that the compartmental models have adopted. In parallel, these models may leverage helpful information from mobility data available from smartphones and other wearable devices. We surveyed both compartmental and contact-based models and presented a contact-based model's application in the simulation of different vaccination strategies.

Key Points

- We present a discussion of models for the analysis of disease spreading
- We discuss advantages and limitations of contact-based models
- We discuss on the possibility to improve the vaccination performance using contact-based models grounded on graph topology considerations
- We formalize a framework for testing vaccination strategies.

Funding

References

1. Humphreys P. Computational models. *Philos Sci* 2002;**69**(S3): S1–11.

2. Tiwary BK. Computational medicine: quantitative modeling of complex diseases. *Brief Bioinform* 2019;**21**(2):429, 01–40.

3. Chen X, Yan CC, Zhang X, *et al*. Drug–target interaction prediction: databases, web servers and computational models. *Brief Bioinform* 2016;**17**(4):696–712.

4. Cirillo D, Valencia A. Big data analytics for personalized medicine. *Curr Opin Biotechnol* 2019;**58**:161–7.

5. Vynnycky E, White R. An introduction to infectious disease modelling. *OUP oxford* 2010.

6. Cannataro M, Guzzi PH, Sarica A. Data mining and life sciences applications on the grid. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2013;**3**(3):216–38.

7. Freifeld CC, Mandl KD, Reis BY, *et al*. Healthmap: global infectious disease monitoring through automated classification and visualization of internet media reports. *J Am Med Inform Assoc* 2008;**15**(2):150–7.

8. Gostin LO. The coronavirus pandemic 1 year on-what went wrong? *JAMA* 2021;**325**(12):1132–3.

9. Guo Y-R, Cao Q-D, Hong Z-S, *et al*. The origin, transmission and clinical therapies on coronavirus disease 2019 (covid-19) outbreak–an update on the status. *Mil Med Res* 2020;**7**(1):1–10.

10. Le TT, Andreadakis Z, Kumar A, *et al*. The covid-19 vaccine development landscape. *Nat Rev Drug Discov* 2020;**19**(5):305–6.

11. Das JK, Tradigo G, Veltri P, *et al*. Data science in unveiling covid-19 pathogenesis and diagnosis: evolutionary origin to drug repurposing. *Brief Bioinform* 2021;**22**(2):855–72.

12. Ortuso F, Mercatelli D, Guzzi PH, *et al*. Structural genetics of circulating variants affecting the sars-cov-2 spike/human ace2 complex. *Journal of Biomolecular Structure and Dynamics* 2021; 1–11.

13. Gonçalves J, da Silva PG, Reis L, *et al*. Surface contamination with sars-cov-2: a systematic review. *Sci Total Environ* 2021;**798**:149231.

14. Gwaltney Jr JM, Owen J, Hendley. Transmission of experimental rhinovirus infection by contaminated surfaces. *Am J Epidemiol* 1982;**116**(5):828–33.

15. Bai Y, Yang B, Lin L, *et al*. Optimizing sentinel surveillance in temporal network epidemiology. *Sci Rep* 2017;**7**(1):1–10.

16. Zhu P, Dai X, Li X, *et al*. Community detection in temporal networks via a spreading process. *EPL (Europhysics Letters)* 2019;**126**(4):48001.

17. Alguliyev R, Aliguliyev R, Yusifov F. Graph modelling for tracking the covid-19 pandemic spread. *Infectious Disease Modelling* 2021;**6**:112–22.

18. Bryant P, Elofsson A. Modelling the dispersion of sars-cov-2 on a dynamic network graphmedRxiv. 2020.

19. Karaivanov A. A social network model of covid-19. *Plos one* 2020;**15**(10):e0240878.

20. Zaplotnik Ž, Gavrić A, Medic L. Simulation of the covid-19 epidemic on the social network of slovenia: Estimating the intrinsic forecast uncertainty. *PloS one* 2020;**15**(8):e0238090.

21. Das JK, Roy S, Guzzi PH. Analyzing host-viral interactome of sars-cov-2 for identifying vulnerable host proteins during covid-19 pathogenesis. *Infect Genet Evol* 2021;104921.

22. Patil R, Dave R, Patel H, *et al*. Assessing the interplay between travel patterns and sars-cov-2 outbreak in realistic urban setting. *Applied Network Science* 2021;**6**(1):1–19.

23. Borgatti SP. Identifying sets of key players in a social network. *Comput Math Org Theor* 2006;**12**:21–34.

24. Paluch R, Xiaoyan L, Suchecki K, *et al*. Fast and accurate detection of spread source in large complex networks. *Sci Rep* 2018;**8**(1): 1–10.

25. Saunders HA, Schwartz J-M. Covid-19 vaccination strategies depend on the underlying network of social interactions. *Sci Rep* 2021;**11**(1):1–10.

26. Ash C. Superspreaders are local and disproportionate. *Science* 2017;**355**(6329):1036–6.

27. Stegehuis C, Van Der Hofstad R, Van Leeuwaarden JSH. Epidemic spreading on complex networks with community structures. *Sci Rep* 2016;**6**(1):1–7.

28. Mazza T, Romanel A, Jordán F. Estimating the divisibility of complex biological networks by sparseness indices. *Brief Bioinform* 2010;**11**(3):364–74.

29. Menniti S, Castagna E, Mazza T. Estimating the global density of graphs by a sparseness index. *Appl Math Comput* 2013;**224**:346–57.

30. Yang W, Chakrabarti D, Wang C, *et al.* Epidemic spreading in real networks: An eigenvalue viewpoint. In: *22nd International Symposium on Reliable Distributed Systems, 2003. Proceedings.* IEEE, 2003, 25–34.

31. Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* 1976;**22**(4):403–34.

32. Nåsell I. The quasi-stationary distribution of the closed endemic sis model. *Advances in Applied Probability* 1996;**28**(3):895–932.

33. Britton T. Stochastic epidemic models: a survey. *Math Biosci* 2010;**225**(1):24–35.

34. Sturniolo S, Waites W, Colbourn T, *et al.* Testing, tracing and isolation in compartmental models. *PLoS Comput Biol* 2021;**17**(3):e1008633.

35. Contreras S, Dehning J, Mohr SB, *et al.* Low case numbers enable long-term stable pandemic control without lockdowns. *Sci Adv* 2021;**7**(41):eabg2243.

36. Contreras S, Dehning J, Loidolt M, *et al.* The challenges of containing sars-cov-2 via test-trace-and-isolate. *Nat Commun* 2021;**12**(1):1–13.

37. Anderson SC, Edwards AM, Yerlanov M, *et al.* Quantifying the impact of covid-19 control measures using a bayesian model of physical distancing. *PLoS Comput Biol* 2020;**16**(12):1–15.

38. Rossa FD, Salzano D, Di Meglio A, *et al.* A network model of italy shows that intermittent regional strategies can alleviate the covid-19 epidemic. *Nat Commun* 2020;**11**(1):1–9.

39. Levy N, Iv M, Yom-Tov E. Modeling influenza-like illnesses through composite compartmental models. *Physica A: Statistical Mechanics and its Applications* 2018;**494**:288–93.

40. Coburn BJ, Wagner BG, Blower S. Modeling influenza epidemics and pandemics: insights into the future of swine flu (h1n1). *BMC Med* 2009;**7**(1):1–8.

41. Ljubic B, Gligorijevic D, Gligorijevic J, *et al.* Social network analysis for better understanding of influenza. *J Biomed Inform* 2019;**93**:103161.

42. Nielsen BF, Sneppen K, Simonsen L, *et al.* Social network heterogeneity is essential for contact tracingmedRxiv. 2020.

43. Giordano G, Blanchini F, Bruno R, *et al.* Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nat Med* 2020;**26**(6):855–60.

44. Köhler J, Schwenkel L, Koch A, *et al.* Robust and optimal predictive control of the covid-19 outbreak. *Annual Reviews in Control* 2021;**51**:525–39.

45. Giordano G, Colaneri M, Di Filippo A, *et al.* Modeling vaccination rollouts, sars-cov-2 variants and the requirement for non-pharmaceutical interventions in italy. *Nat Med* 2021;**27**(6):993–8.

46. Oliveira JF, Jorge DCP, Veiga RV, *et al.* Mathematical modeling of covid-19 in 14.8 million individuals in bahia, brazil. *Nat Commun* 2021;**12**(1):1–13.

47. Keeling MJ, Grenfell BT. Individual-based perspectives on r0. *J Theor Biol* 2000;**203**(1):51–61.

48. Tracy M, Cerdá M, Keyes KM. Agent-based modeling in public health: current applications and future directions. *Annu Rev Public Health* 2018;**39**:77–94.

49. Hoya White S, Martín Del Rey A, Rodríguez G, *et al.* Modeling epidemics using cellular automata. *Appl Math Comput* 2007;**186**(1):193–202.

50. Sharma N, Gupta AK. Impact of time delay on the dynamics of seir epidemic model using cellular automata. *Physica A: Statistical Mechanics and its Applications* 2017;**471**:114–25.

51. Sun K, Wang W, Gao L, *et al.* Transmission heterogeneities, kinetics, and controllability of sars-cov-2. *Science* 2021;**371** (6526).

52. Moein S, Nickaeen N, Roointan A, *et al.* Inefficiency of sir models in forecasting covid-19 epidemic: a case study of isfahan. *Sci Rep* 2021;**11**(1):1–9.

53. Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. *Phys Rev Lett* 2001;**86**(14):3200.

54. Ajelli M, Gonçalves B, Balcan D, *et al.* Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infect Dis* 2010;**10**(1):1–13.

55. Liu Q-H, Ajelli M, Aleta A, *et al.* Measurability of the epidemic reproduction number in data-driven contact networks. *Proc Natl Acad Sci* 2018;**115**(50):12680–5.

56. Koher A, Lentz HHK, Hövel P, *et al.* Infections on temporal networks-a matrix-based approach. *PloS one* 2016;**11**(4):e0151209.

57. Darbon A, Valdano E, Poletto C, *et al.* Network-based assessment of the vulnerability of italian regions to bovine brucellosis. *Prev Vet Med* 2018;**158**:25–34.

58. Marziano V, Poletti P, Béraud G, *et al.* Modeling the impact of changes in day-care contact patterns on the dynamics of varicella transmission in france between 1991 and 2015. *PLoS Comput Biol* 2018;**14**(8):e1006334.

59. Chang S, Pierson E, Koh PW, *et al.* Mobility network models of covid-19 explain inequities and inform reopening. *Nature* 2021;**589**(7840):82–7.

60. Hâncean M-G, Slavinec M, Perc M. The impact of human mobility networks on the global spread of covid-19. *Journal of Complex Networks* 2020;**8**(6):cnaa041.

61. Hartnett GS, Parker E, Gulden TR, *et al.* Modelling the impact of social distancing and targeted vaccination on the spread of covid-19 through a real city-scale contact network. *Journal of Complex Networks* 2021;**9**(6):cnab042.

62. Peixoto PS, Marcondes D, Peixoto C, *et al.* Modeling future spread of infections via mobile geolocation data and population dynamics. an application to covid-19 in brazil. *PloS one* 2020;**15**(7):e0235732.

63. Müller SA, Balmer M, Charlton W, *et al.* Predicting the effects of covid-19 related interventions in urban settings by combining activity-based modelling, agent-based simulation, and mobile phone data. *PLoS ONE* 2021;**16**(10):e0259037.

64. Guzmán-Merino M, Durán C, Marinescu M-C, *et al.* Assessing population-sampling strategies for reducing the covid-19 incidence. *Comput Biol Med* 2021;**139**:104938.

65. Kerr CC, Stuart RM, Mistry D, *et al.* Covasim: an agent-based model of covid-19 dynamics and interventions. *PLoS Comput Biol* 2021;**17**(7):e1009149.

66. Kucharski AJ, Klepac P, Conlan AJK, *et al.* Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of sars-cov-2 in different settings: a

mathematical modelling study. *Lancet Infect Dis* 2020;**20**(10): 1151–60.

67. Cooper I, Mondal A, Antonopoulos CG. A sir model assumption for the spread of covid-19 in different communities. *Chaos, Solitons & Fractals* 2020;**139**:110057.

68. Gnanvi J, Salako KV, Kotanmi B, *et al*. On the reliability of predictions on covid-19 dynamics: A systematic and critical review of modelling techniques. *Infectious Disease Modelling* 2021.

69. Dashtbali M, Mirzaie M. A compartmental model that predicts the effect of social distancing and vaccination on controlling covid-19. *Sci Rep* 2021;**11**(1):1–11.

70. Zhang J, Litvinova M, Wang W, *et al*. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside hubei province, china: a descriptive and modelling study. *Lancet Infect Dis* 2020;**20**(7):793–802.

71. Parino F, Zino L, Calafiore GC, *et al*. A model predictive control approach to optimally devise a two-dose vaccination rollout: A case study on covid-19 in italy. *International Journal of Robust and Nonlinear Control* 2021.

72. Antonini C, Calandrini S, Bianconi F. A modeling study on vaccination and spread of sars-cov-2 variants in italy. *Vaccine* 2021;**9**(8):915.

73. Viana J, van Dorp CH, Nunes A, *et al*. Controlling the pandemic during the sars-cov-2 vaccination rollout. *Nat Commun* 2021;**12**(1):1–15.

74. Esteban EP, Almodovar-Abreu L. Assessing the impact of vaccination in a covid-19 compartmental model. *Informatics in Medicine Unlocked* 2021;100795.

75. Gardner BJ, Kilpatrick AM. Contact tracing efficiency, transmission heterogeneity, and accelerating covid-19 epidemics. *PLoS Comput Biol* 2021;**17**(6):e1009122.

76. Newman ME, Barabási A-LE, Watts DJ. *The structure and dynamics of networks*. Princeton university press, 2006.

77. Bubar KM, Reinholt K, Kissler SM, *et al*. Model-informed covid-19 vaccine prioritization strategies by age and serostatus. *Science* 2021;**371**(6532):916–21.

78. Khamsi R. If a coronavirus vaccine arrives, can the world make enough. *Nature* 2020;**580**(7805):578–80.

79. Galicia JC, Guzzi PH, Giorgi FM, *et al*. Predicting the response of the dental pulp to sars-cov2 infection: a transcriptome-wide effect cross-analysis. *Genes & Immunity* 2020;**21**(5):360–3.

80. Guzzi PH, Mercatelli D, Ceraolo C, *et al*. Master regulator analysis of the sars-cov-2/human interactome. *J Clin Med* 2020;**9**(4):982.

81. Cannistraci CV, Capua I. Age-adaptive and gender-balanced social interventions might be stepping stones to the next 'new normal' in covid-19 outbreak. 2020.

82. Goldstein JR, Cassidy T, Wachter KW. Vaccinating the oldest against covid-19 saves both the most lives and most years of life. *Proc Natl Acad Sci* 2021;**118**(11).

83. Buckner JH, Chowell G, Springborn MR. Dynamic prioritization of covid-19 vaccines when social distancing is limited for essential workers. *Proc Natl Acad Sci* 2021;**118**(16).

84. Maheshwari P, Albert R. Network model and analysis of the spread of covid-19 with social distancing. *Applied network science* 2020;**5**(1):1–13.

85. Jentsch PC, Anand M, Bauch CT. Prioritising covid-19 vaccination in changing social and epidemiological landscapes: a mathematical modelling study. *Lancet Infect Dis* 2021.

86. Freeman L. Centrality in social networks: Conceptual clarification. *Social Networks* 1979;**1**(3):215–39.

87. Everett MG, Borgatti SP. The centrality of groups and classes. *The Journal of Mathematical Sociology* 1999;**23**(3):181–201.

88. Shannon P, Markiel A, Ozier O, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**(11):2498–504.

89. Parca L, Truglio M, Biagini T, *et al*. Pyntacle: a parallel computing-enabled framework for large-scale network biology analysis. *GigaScience* 2020;**9**(10):giaa115.

90. Miller JC, Ting T. Eon (epidemics on networks): a fast, flexible python package for simulation, analytic approximation, and analysis of epidemics on networks. *Journal of Open Source Software* 2019;**4**(44):1731.