Geospatially Referenced Demographic Agent-Based Modeling of SARS-CoV-2-Infection (COVID-19) Dynamics and Mitigation Effects in a Real-world Community

Adler SO[1], Bodeit O[1,2], Bonn L[1], Goldenbogen B[1], Escalera-Fanjul X[1], Haffner J[1], Karnetzki M[1], Korman A[1], Krantz M[1], Linding R[1,3], Maintz I[1], Mallis L[1], Moran Torres RU[1], Prawitz H[1], Seeger M[1], Segelitz P[1], Wodke JAH[1] and Klipp E[1,*]
alphabetical order of authors

[1] Humboldt-Universität zu Berlin, Theoretical Biophysics, Invalidenstr. 42, 10115 Berlin, Germany
[2] Institut für Biochemie, Charité – Universitätsmedizin Berlin, Berlin, Germany
[3] Rewire Tx, Humboldt-Universität zu Berlin
[*] Correspondence to: edda.klipp@hu-berlin.de

Abstract

Re-opening societies and economies across the globe following the initial wave of the severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) pandemic requires scientifically-guided decision processes and policy development. Public health authorities now consider it highly likely that transmission of SARS-CoV-2 and COVID-19 will follow a pattern of seasonal circulation globally. To guide mitigation strategies and tactics in a location-specific manner, accurate simulation of prolonged or intermittent patterns of social/physical distancing is required in order to prevent healthcare systems and communities from collapsing. It is equally important to capture the stochastic appearance of individual transmission events. Traditional epidemiological/statistical models cannot make predictions in a geospatial temporal manner based on human individuals in a community. Thus, the challenge is to conduct spatio-temporal simulations of transmission chains with real-world geospatial and georeferenced information of the dynamics of the disease and the effect of different mitigation strategies such as isolation of infected individuals or location closures. Here, we present a stochastic, geospatially referenced and demography-specific agent-based model with agents representing human beings and include information on age, household composition, daily occupation and schedule, risk factors, and other relevant properties. Physical encounters between humans are modeled in a time-dependent georeferenced network of the population. The model (GERDA-1) can predict infection dynamics under normal conditions and test the effect of different mitigation scenarios such as school closures, reduced social contacts as well as closure or reopening of public/work spaces. Specifically, it also includes the fate and influence of health care workers and their access to protective gear. Key predictions so far entail:

(i) the effect of specific groups on the spreading, specifically that children in school contribute substantially to distribution.
(ii) the result of reopening society depends crucially on how strict the measures have been during lock-down.
(iii) the outcome of reopening is a stochastic process - in the majority of cases, we must expect a second wave, in some cases not. To the best of our best knowledge, the GERDA-1 model is the first model able to predict a bimodal behavior of SARS-Cov-2 infection dynamics.

Given the criticality of the global situation, informing the scientific community, decision makers and the general public seems prudent. Therefore, we here provide a pre-print of the GERDA-1 model together with a first set of predictions and analyses as work in progress.

Introduction

As SARS-CoV-2 infection is spreading around the world it is inferring multi-dimensional damage to humanity: millions of COVID-19 patients are bringing healthcare systems close to collapse, halting or suppressing global and local economies, and normal human activity. In response, countries and communities are scrambling to fight the virus with a series of different measures and strategies aimed at preventing new infections whilst providing optimal treatment of patients and aiming to re-open economies as swiftly as possible.

To assist policy makers and governments in choosing different exit/re-entry strategies, data-driven mathematical and computational modelling can predict the path and severity of infection, the expected number of fatalities and the effect of different mitigation measures.
Considering the lack of information about not only the disease itself but also its near- and long-term impact on healthcare systems and society, models are required to reliably predict the effect of alternative strategies to both “re-open” societies and allow the general population to return to work and social activities. Therefore, the scientific community has a great responsibility to create such mathematical models and ensure that the predictions are as precise and adapted to reality as possible, as the accuracy of the predictions can directly influence society.
In short, there are chiefly two different computational modeling approaches used to predict the spread of a pathogen. On the one hand, deterministic models often use ordinary differential equations (ODEs) to depict trajectories of defined variables that change over time and are dependent on population-level metrics, data and statistics. On the other hand, stochastic models, such as agent-based models (ABM), use a set of rules and probabilities in order to define the behavior of agents, in this case individuals, which can be tracked over time in order to extract population-level metrics that can always be traced back to their agent-level origin. In typical epidemiological models different groups of people can be considered based on common features (e.g. based on demographics, risk groups, age etc.), where particularly the group of susceptibles (S), the group of infected (I) and the group of recovered (R) individuals are usually of great interest [1]. Recovered individuals are assumed

to have acquired immunity to some degree (though this is still an open question for SARS-CoV-2). The transition between these states depends on **(i)** a basic reproduction number R_0 describing how likely it is that infected individuals transmit the infection to susceptible individuals and **(ii)** parameters describing the recovery. More advanced models include further relevant states such as exposed (E), hospitalized (H) or being treated in an ICU, as recently presented by an ODE model of the Neher laboratory [see https://covid19-scenarios.org/]. An example for a respective ABM model has already been implemented in the tool Netlogo [3].

Here we present an elaborate model (GERDA-1) that can provide simulated input for informed decision processes. To this end, we developed an agent-based model that incorporates the accepted different stages of SARS-Cov-2 infection, COVID-19 disease and recovery that takes demographic data, realistic daily scenarios and, importantly, the *physical location* of individuals into consideration. GERDA-1 can predict the effects of different tailored scenarios and measures for mitigation depending on the location (e.g. workplace, school, public places such as shopping malls, etc) but also dependending on the actual time of day. To this end, we have integrated a large amount of publicly available data, e.g., on locations, age distribution, geographical information, and sociological data for typical numbers and types of social contacts in the German population.
Most importantly, we have taken into account the distinct role of front-line health care workers, medical doctors and first-responders, because these groups are most exposed to the virus and therefore have an increased risk of contracting it. It is emerging that SARS-Cov-2 is inducing the disease in a dose-dependent manner, a characteristic that is particularly problematic for healthcare providers. Consequently, it is critical to model the burden and surge-requirements for hospitals/medical systems, as this very much depends on essential staff to ensure the functioning of the medical system including local doctor offices and hospitals.

Results

Given the critical importance of conducting location- and situation-specific predictions of virus spreading and optimal measures to suppress the infection rate to a range that can be handled by the healthcare system. Therefore, GERDA-1 represents the virus propagation and effects of mitigation measures within a concrete (German) population. The model tracks the number of individuals with specific infection states at a specific physical location through time. Because the model is georeferenced, it makes use of the concrete number of residential buildings, workplaces, schools or public places in a given community/municipality as the space for potential infection chain initiation by human contact. The approach is displayed in **Figure 1**. GERDA-1 is based on demographic data from the German census with respect to age distributions, household sizes and composition. As an example, we used data from the municipality Gangelt (**Figure 2A** for a map), a place that experienced a widely

noticed COVID-19 outbreak in Germany. Gangelt had more than 300 confirmed cases among 12,000 inhabitants with the outbreak after a carnival event in March 2020 followed by widespread transmission within the community.

It is important to note that a model remains an abstraction and it is neither repetition of history nor can it be traced back to real individuals, but allows to study realistic instead of real scenarios whilst safe-guarding privacy and anonymity for members of the public

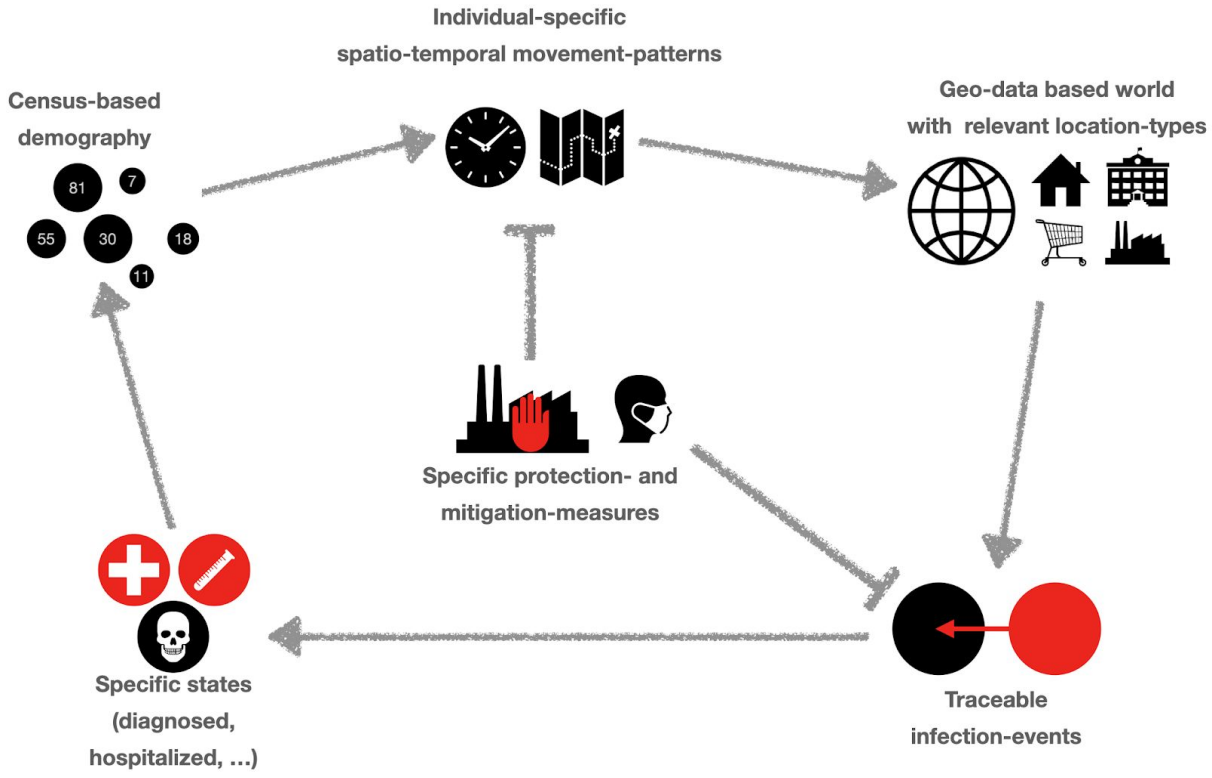


Figure 1: **Overview of the GERDA-1 model** Agents representing individuals are drawn from distributions representing census-based demography. They have individual properties, such as age, risk, infection state and follow a specific weekly schedule. Each day, an individual visits locations in their geo-data-based world according to their own schedule. Individuals can with a certain likelihood either become infected or infect other individuals in these locations. Following an infection an individual can be diagnosed, admitted to hospital and/or ICU, recover or succumb to the disease. Mitigation measures influence the likelihood for infections, e.g., by excluding (closing down of) specific types of locations or by wearing personal protective equipment (PPE).

The GERDA-1 Model

In the model each human being is described as an autonomous agent who is always present at one specific physical location at a given time point; thus even deceased individuals are assigned to a morgue location. In the current basic version (v.1.0), the location can be one of the following types: residential buildings (home), work places, schools, hospitals, and public places. These *locations* are initialized automatically from openly available data for a given

municipality using geographical data from OpenStreetMap. Here, we used Gangelt as an example (see **Figure 2A**).

World initialization: Each building is considered a location as long as its floor area is large enough and it is not a non-residential building, e.g. a windmill. The building types are assigned according to labels in the input data. For example buildings marked as 'office' or 'industrial' are assigned as workplaces while buildings marked as 'public' or 'church' are assigned as public places. All buildings that are not labelled specifically are assigned as residential buildings or homes. If the input data did not contain a location for a hospital or morgue, these locations were added artificially at the margin of the world, during the initialization process (see **Figure 2D**).

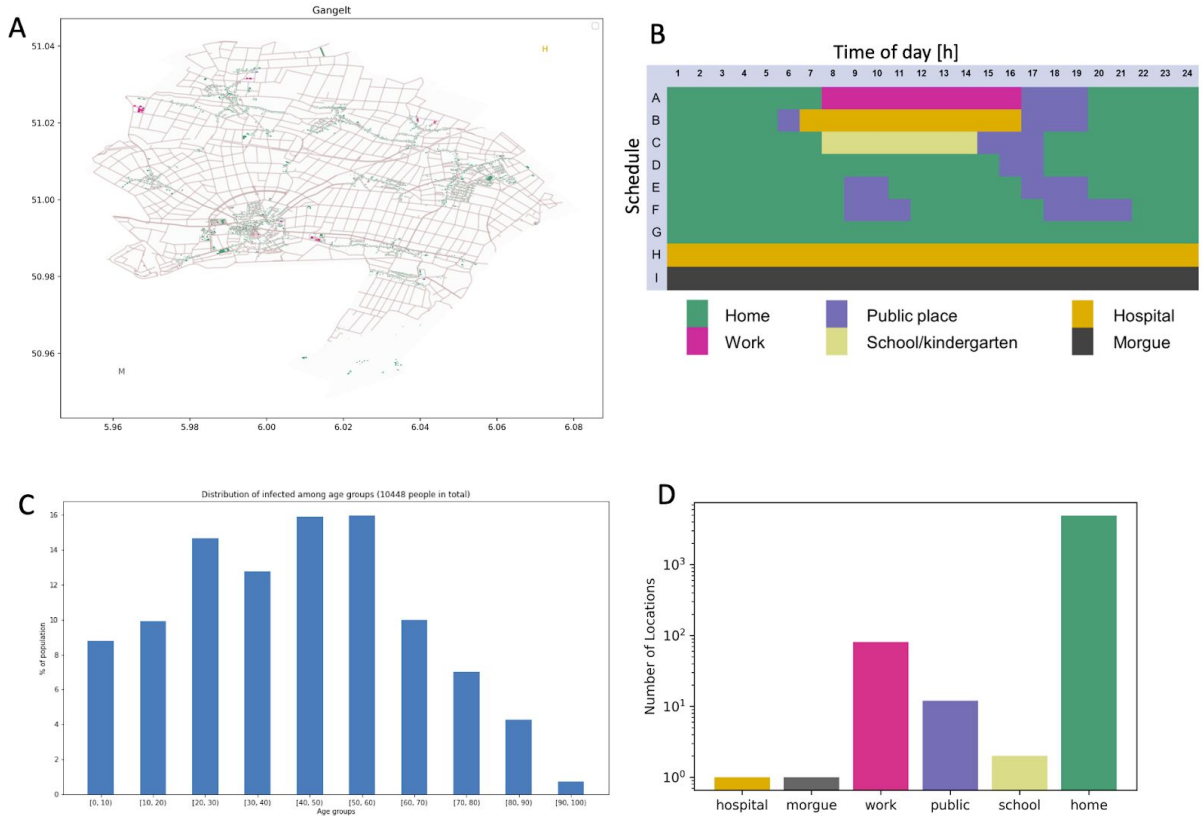


Figure 2. **Georeference of the model. A**) Map of Gangelt as defined by OpenStreetMap. Overlaid are indicators for the location type assigned to each building. Hospital (H) and morgue (M) is introduced (in case they are missing) at the border of the map. **B**) Exemplary schedules for individuals (A - Adult/employee, weekdays, no mitigation; B - Medical care worker, work day, no mitigation; C - Student/child, weekdays, no mitigation; D - Student/child, weekdays, schools closed; E - Adult/pensionist, all days, no mitigation; F - Student/child, adult/employee, weekend, no mitigation; G - Diagnosed person, all days; H - Hospitalized person, all days; I - Deceased person - all days). **C**) Resulting age distribution of the modelled population. **D**) Number of locations per type. Colors in A, B, and D correspond to the same types of location, respectively.

Initializing the agent population: Each residential building is home to one household that was randomly drawn from a distribution, based on German census data. According to the respective household type and considering the reported demography, agents with adequate ages are randomly generated and given different properties such as profession, risk state

and infection status (see **Figure 2C**). Additionally each agent is assigned a specific weekly schedule, that comprises times spent at home, work, public places etc., which are based on predefined flexible schedules for different age groups and types of individuals (see **Figure 2B**). These can differ from individual to individual and for different days of the week, e.g. the weekend.
This spatio-temporal network, defined by periodically recurring movement patterns, constitutes the environment in which the agents interact and the infection spreads. Despite some degree of freedom; most agents operate in defined sub-networks, specified by regularly visited locations.

During the initialization of a modeled world the infection status of all agents is set to susceptible (**S**). At the beginning of a simulation the status of a predefined number of agents is set to infected (I), chosen from a minimal number of households. During each time-step of the simulation, the agents visit the locations specified by their respective schedules and the infection spreads across the spatio-temporal network of interacting agents.

Characterization of agents: Agents are characterized according to their health status as susceptible (S), infected (I), recovered (R) or deceased (D). Agents among the group of infected individuals (I) obtain sub-states specifying their condition as (only) infected (I), diagnosed (I^d), hospitalized (I^d_H), or being in an ICU (I^d_{ICU}) (see **Figure 3A**). These states and sub-states are accompanied by corresponding schedules. Upon hospitalization (and during stay in the ICU) the agent follows a hospital-schedule, which places it in the hospital until recovery or death (in the basic model we do not consider possible re-infection following release from hospital). Diagnosed individuals are considered to be quarantined in their respective homes. While the recovery of an agent introduces the reacquisition of its regular schedule, its death results in assignment of the deceased-schedule (see **Figure 3D**).

GERDA-1 specifically represents the group of healthcare workers and medical professionals by defining corresponding schedules with medical facilities as workplaces. The high exposure to infected individuals in this location yields an elevated risk for infection in agreement with hitherto observation made for SARS-Cov-2 in multiple outbreak theaters. In addition, the risk for medical professionals under different levels of personal protection measures can be assessed in isolation by targeted data analysis.

Modeling the spreading of SARS-Cov-2 infection: Each agent is considered to encounter (unilateral) interactions with other agents, according to a specified interaction-frequency (one per hour), based on the underlying georeferenced network and the agents respective schedules (**Figure 3C**). The interaction partner is randomly picked among all other agents present at the same location at the same time. Given a susceptible agent (S) has been assigned to an infected agent (I) as interaction partner, the infection transmission is a situation-dependent property occurring with a specific probability, $P_{infection}$, upon which

the uninfected agent's state might transit from susceptible to infected (**Figure 3B**). This infection probability reads

$$P_{infection} = P_{infection}(t) = k_I \cdot I^I(t) \cdot S^S$$

where I^I is the infectivity of an infected agent (I) depending on its duration of infection and S^S the susceptibility of a susceptible agent (S). k_I is the infection rate, which represents individual protective measures of both parties (e.g., wearing face-masks) or location-specific factors such as the more strict hygiene regime in medical facilities.

State transitions of infected agents: Infected agents have a probability (per hour) to change their status to either (R) or (D) or change their sub-status to (I^d), (I^d_H) or (I^d_{ICU}). These probabilities are derived from published rates for those processes given in the **Supplementary Material** (**Table S2**). The transition probabilities that we used to simulate the model are calculated for all transitions as probability per hour, where their dependence on age and the duration of ongoing (sub-)states is taken into account.

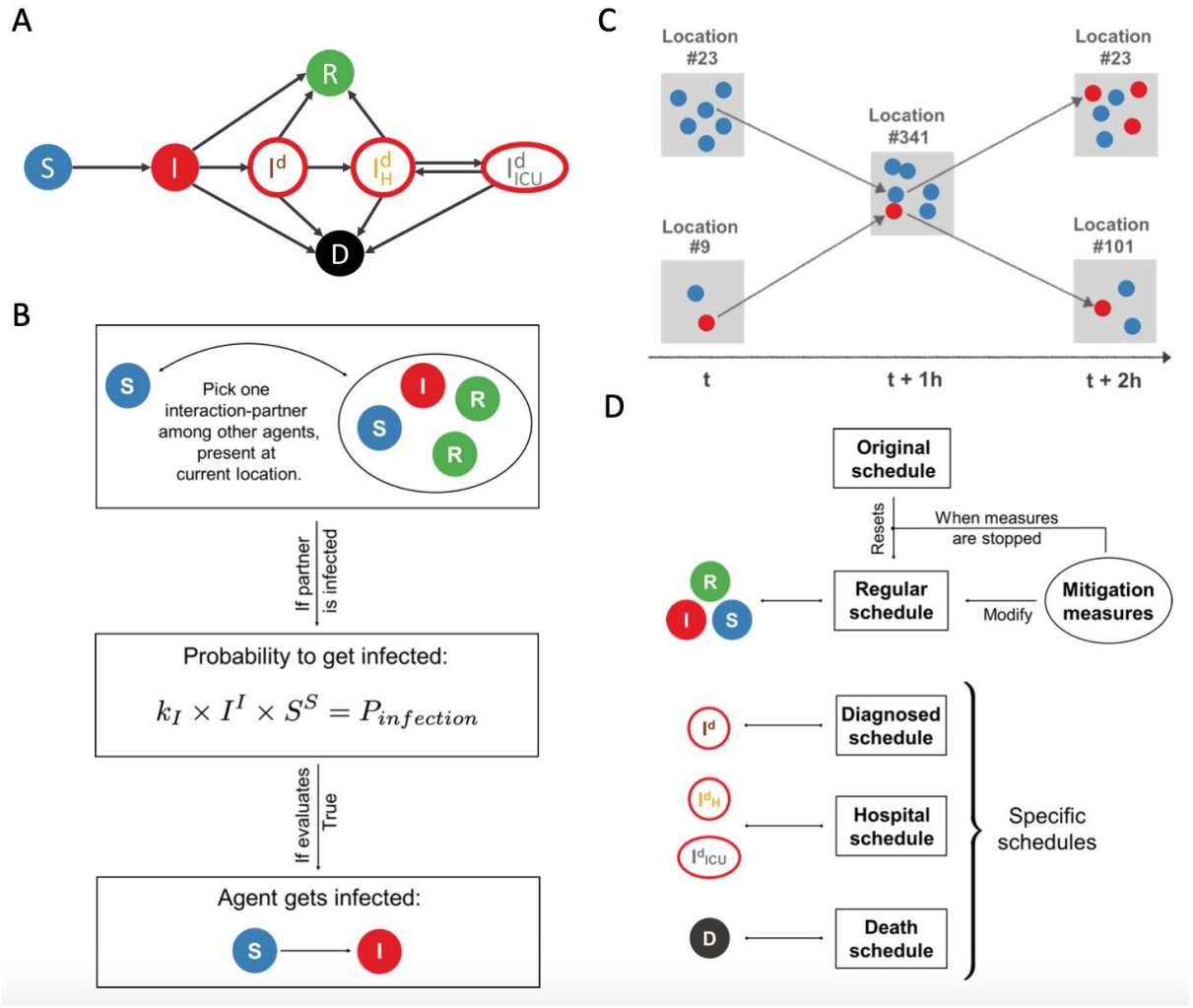


Figure 3. **Principles of agents/individuals, their states and schedules. A**) For state transitions in the model, we consider a transition from susceptible (S) to infected (I), based on interaction between (S) and (I). Infected individuals comprise those not diagnosed (I in full circle) and those diagnosed (I^d), hospitalized (I^d_H) and those in ICU (I^d_{ICU}). All infected individuals have a probability of dying (D), patients from the ICU can return to the hospital (non-ICU or recovery ward), and all infected except ICU patients can recover (R). The colors correspond to the colors used below in time course plots. **B**) Infection process and probabilities, as an example for status transitions in the model (I^I : infectivity of an infected agent; S^S: susceptibility of a susceptible agent ; k_I: infection rate). **C**) Schematic representation of infection spreading via interactions at locations over time. At time *t*, a susceptible and an infected agent move to a location, specified by their respective schedules. They encounter each other at time *t+1* and the infection can occur; subsequently they move to other scheduled locations, where they may spread the infection further. **D**) Overview of different schedules. Agents with states (S), (R) and (I) follow their regular schedules, which are modulated by the restrictions imposed by mitigation-measures. Upon lifting of these measures, the schedules are reset to their original state. Diagnosed, deceased and hospitalized agents follow specific schedules.

Agent data from simulation: All locations and (sub-)states of individual agents are recorded and their time-course over the entire simulation is stored to be used as input for subsequent simulations and for analyses and plotting. Furthermore, information on the infection network (when, where and by whom every modelled agent was infected) can be extracted once a simulation is completed. To test the effect of different mitigation

measures, parameter values, or the application and relaxation of social distancing, repeated simulations (with the different measures imposed) can be performed with identical initializations or starting from a selected time point of previous simulations to compare alternative trajectories of disease-spread.

Simulation Results of the Model - Baseline Version (GERDA-1)

We start the model with a basic initialization of households, workplaces, public places, schools, a hospital, and a morgue as well as a total number of infected individuals. The total number of individuals originates from assigning a household (drawn from the demographic distributions) to each residential building. The baseline model is key to subsequent evaluation of the predicted outcome of the outbreak without any implemented measures such as social distancing or home offices and closure of schools (unlimited and silent community spread). The dynamical behavior of the population and their stages is displayed in **Figure 4**.

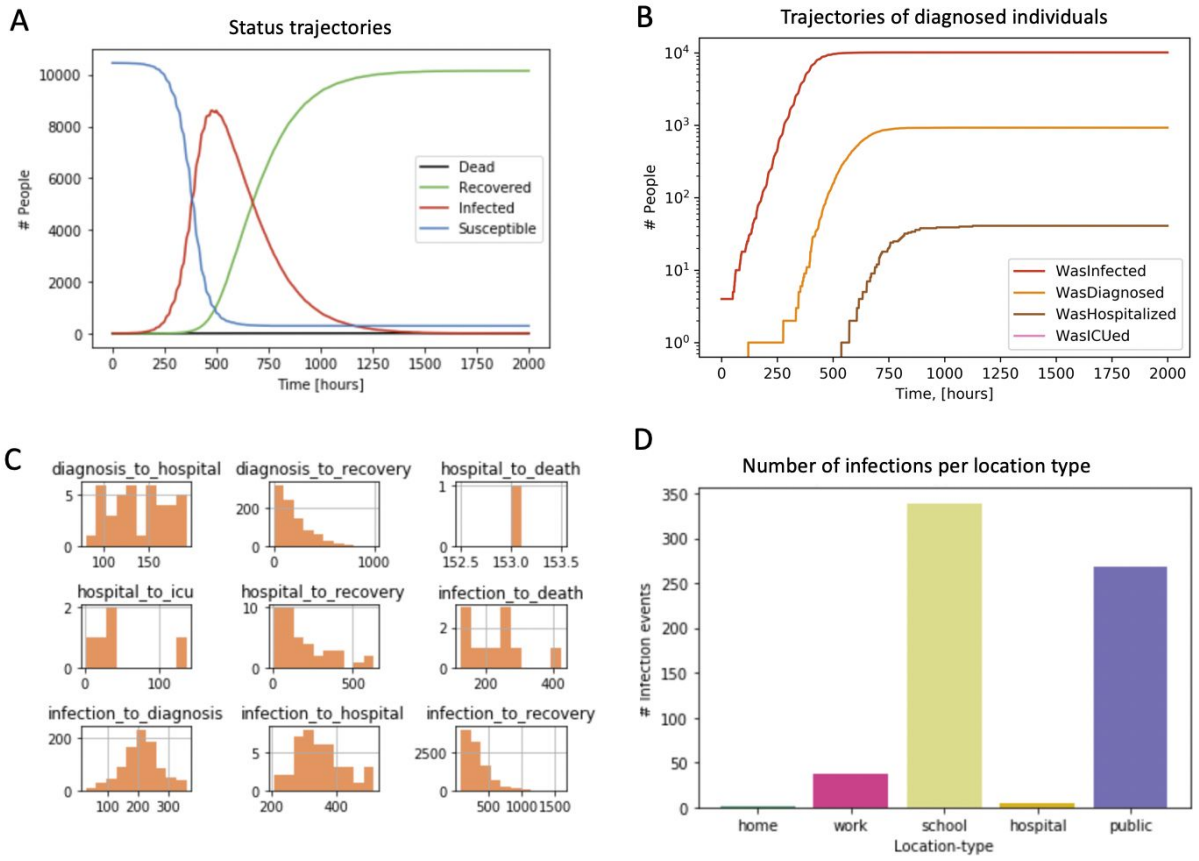


Figure 4. **Simulation results for the baseline scenario for the modeled municipality Gangelt.** Here, we started with 5 infected individuals. **A**) Time courses for susceptible (S), infected (I), recovered (R), and deceased (D) individuals. **B**) Time courses for diagnosed (I^d), hospitalized (I^d_H) individuals and those in ICU (I^d_{ICU}). **C**) Distribution of transition times between stati for infected individuals. **D**) Number of infections per location type normalized to the number of locations of each type. Parameter values: Infectivity = 0.5

The dynamics at the different locations are represented in **Figure 5**. Representations for groups of people with different states (S), (I), (R), and (D) at the locations home, school, hospital, public, and work, are in each case summarized across all locations belonging to a certain type (e.g. for all schools together). It can be observed how the wave of infection in the event of no mitigation spreads through the population. It is important to note that the weekly structure of schedules (work or school from Monday to Friday, more visits of public places during the weekends) directly influence the resulting dynamics. Also represented is the morgue as a location for deceased individuals. Age-dependent trajectories are shown in **Supplementary Figure S1**.

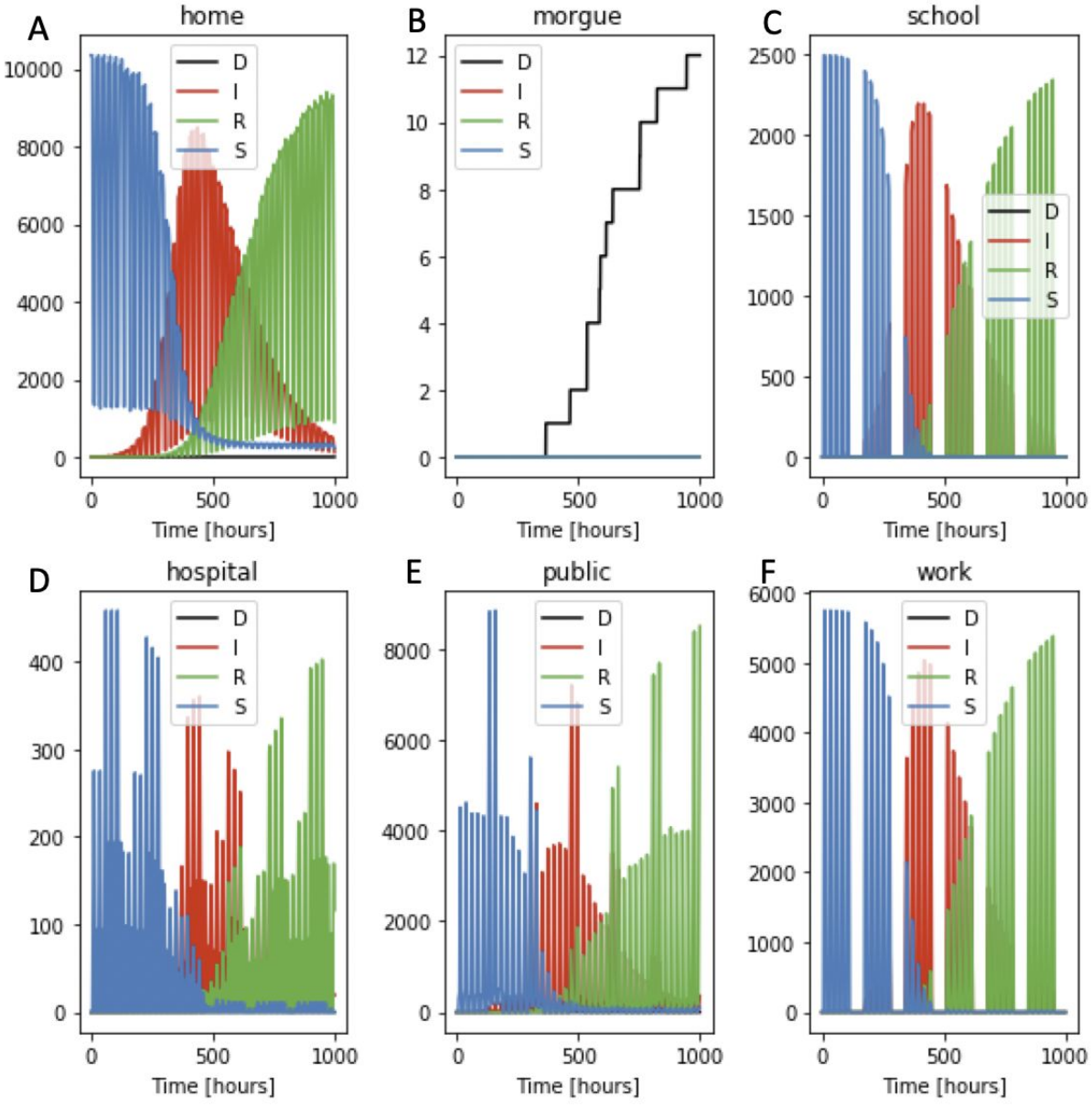


Figure 5. **Exemplary representation of daily occupations of the agents during the outbreak.** Simulation of the same scenario as in Figure 4. **A**) Home represents the residential building where agents spend most of their time. Agents can be either susceptible, infected or recovered. **B**) The morgue represents the location for deceased individuals. **C**), **E**), **F**) Time courses for schools, public places and workplaces. All three types of location represent the dynamics of the infection wave with initially high number of susceptible and later high

number of recovered individuals. School and workplaces also reflect the effect of weekly schedules of the individuals with 5 days mainly at work or school and two days prevalently at home or public places. **D**) The hospital comprised both hospitalized individuals and doctors/nurses who work in the hospital (red susceptible, green recovered).

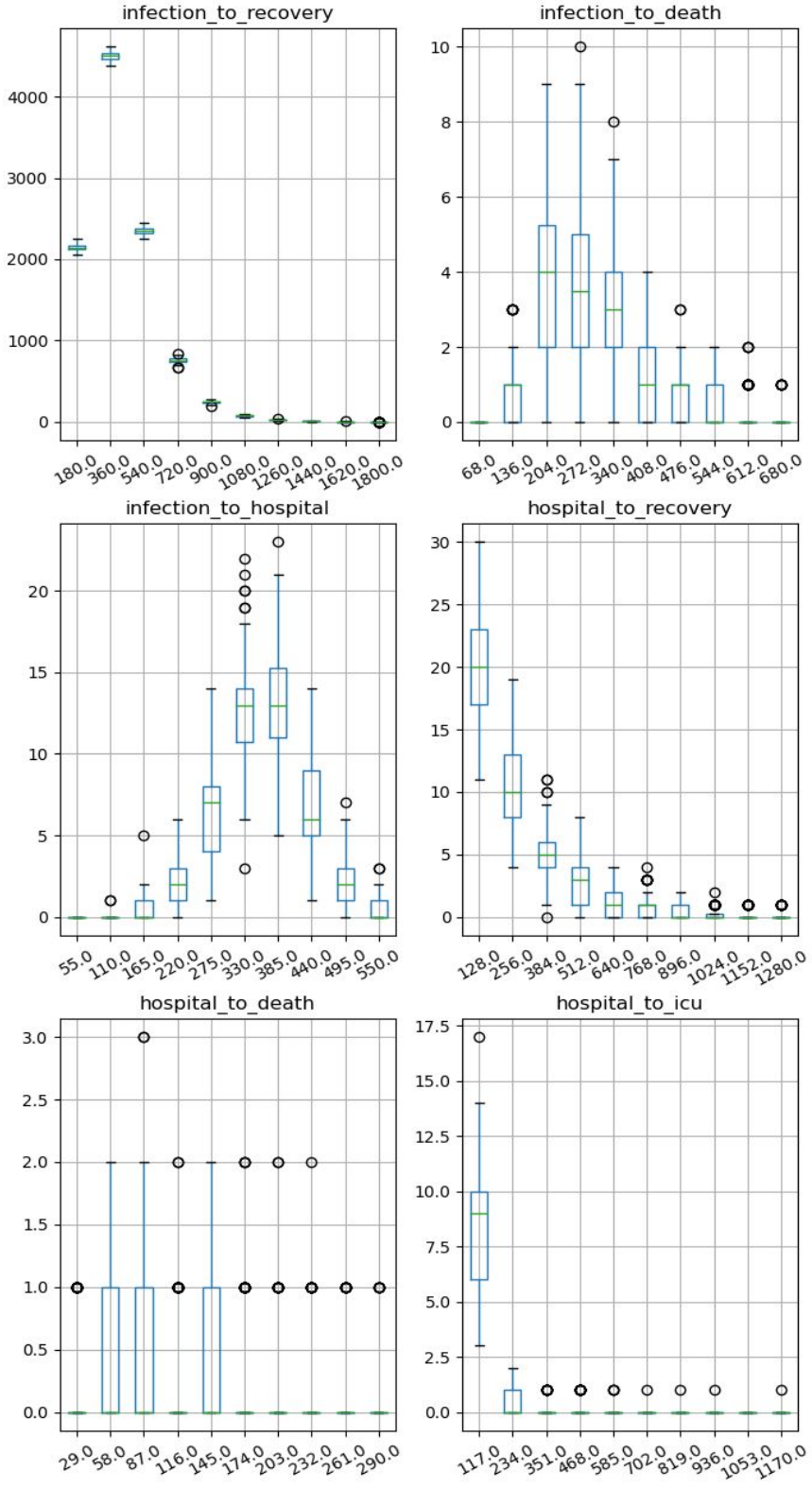


Figure 6. **Resulting distributions, over 100 replicate simulations, of the duration (hours), spent in different states prior to the transition to other specific states.** Different transitions considered in this figure, (starting from top-left I⟶R : infection to recovery , I⟶D : infection to death, I⟶I^d_H : Infection to hospitalization , I^d_H⟶ R : admission to hospital to recovery , I^d_H⟶D : admission to hospital to death , I^d_H⟶ I^d_{ICU} : admission to hospital to onset of ICU-treatment).

Parameters

The model parameters are based on published data for transition periods between infection and diagnosis, diagnosis and hospitalization, frequencies of need for ICU as well as death rates in the German cohort [RKI]. In order to verify that the model predictions reflect these underlying data in a satisfactory manner, we determined distributions of transition frequencies per unit of time for the different state transitions considered in our model (**Figure 6**). These frequencies agree well with the frequencies reported in literature, although not all necessary values for all age groups have been published yet.

The dynamics of infection spreading resulting from the simulation of the baseline scenario is visualized in Movie 1 [https://thbpcloud.de/index.php/s/2lJ8McKxPc0WrvT]. It is important to note that the daily rhythm shown in the movie, results from individuals moving between their respective homes and workplaces. To illustrate this three still images of this movie are shown in **Figure 7**. The infection starts with two infected individuals in one household at time 0 h. At times 100 h and 200 h more and more infected agents are observed, especially at the geographical hubs (e.g., center of the town).

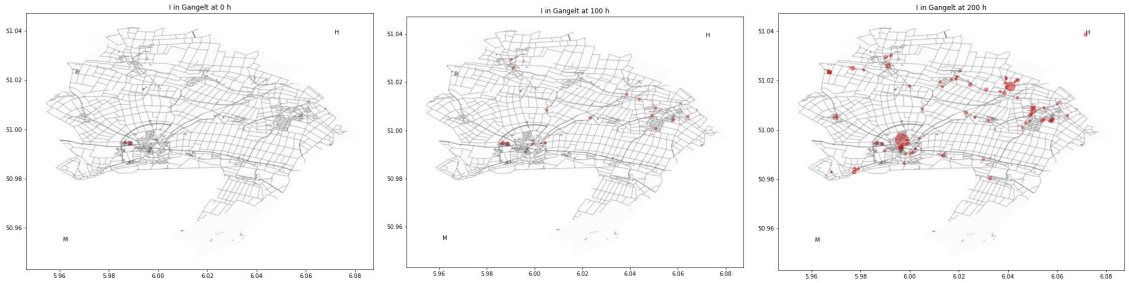


Figure 7. **Still images from Movie 1**. Representing the dynamics of infection spread over the underlying geographic network. Movie1 is available at https://thbpcloud.de/index.php/s/2lJ8McKxPc0WrvT.

Scenarios of Mitigation and Elevation of Contact Restrictions

In order to analyze the effect of mitigation measures and interventions such as lockdowns, or contact prohibition, we tested the following mitigation scenarios:

1. Scaling of the infectivity as proxy for wearing face masks or keeping physical distance
2. Selective closure of public spaces such as schools, general public spaces, workplaces or a combination of these
3. Variation of the infection probability for healthcare workers to simulate limited access to PPE

The effect of the scenario mimicking increased level of social (physical) distancing represented by reduction of infection rate is represented in **Figure 8**. Here, we have compared the baseline version of the model with infectivity values of 0.5 (indicating typical behavior) and 0.3 (corresponding to low social distancing, i.e., the probability to get infected

when meeting an infected individual is decreased by 40%). As a result of the reduced infectivity, the infection wave runs through the population more slowly, i.e., the peak of (I) is lower (about 7000 individuals compared to 8000 at the higher rate) and also occurs later. At the end, fewer people have been infected and moved to the recovered state (about 9500 compared to about 10000 individuals). The reason for the lower number of infected and, therefore, recovered individuals is that in the case of a slower spread of the virus through the population, the overshoot in infected individuals is reduced [5]. Further results for infection rate 0.3 are represented in **Supplementary Figures S2**.

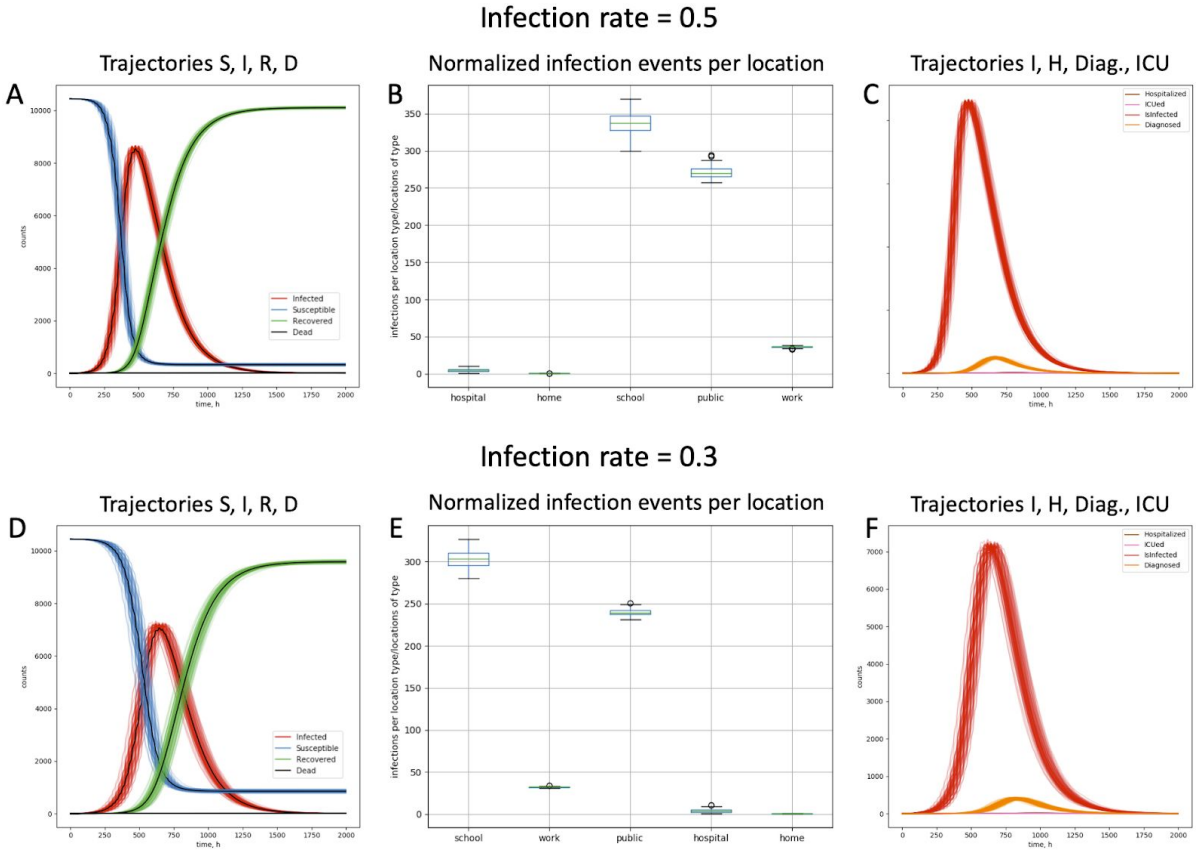


Figure 8. Variation of the infection rate. Upper panel: standard infection rate of 0.5. Lower panel: reduced infection rate of 0.3 mimicking stricter behavior/distancing. **A**) and **D**): Dynamics of S, I, R and D. **B**) and **E**): Number of infections per location (normalized to the number of locations per type). **C**) and **F**): Dynamics of infected, diagnosed, hospitalized and individuals in ICU.

To evaluate the influence of more strict measures to influence the course of the epidemic, we have simulated a series of potential scenarios for lockdown and release of lockdown measures. This includes closures of all locations except of homes and hospitals (i.e., workplaces, schools and public spaces) as well as closure of selected locations. We further tested the selective opening of these locations. **Figure 9** shows exemplary results for closures at 200h (ca. 8 days) and openings at 500h (ca. 20 days) with 100 simulations per scenario. It can be observed that closing of all locations leads to a significant reduction of the infection peak from about 8500 individuals to about 1700 individuals. By only closing public places and schools or public- and workplaces a peak of about 6200 or 3700 infected

individuals, respectively is observed. Thus, closing all locations (except for home and hospital) has a strong advantage compared to closing only selected locations.
Reopening of any selected location type such as schools (**Figure 9D**), work (**Figure 9E**), or all locations (**Figure 9F**) results in a unique phenomenon, namely a bimodal behavior of the system. In this case, the infection declines in some simulations, while reaching a strong second peak of infection in other simulations. This is due to the stochastic nature of the process: the precise behavior of individuals is unpredictable, despite being regulated (even in Germany). The ratio of cases of declining and recurring infection waves varies between the scenarios, as well as the peak of the second infection wave with the strongest peak when all public activities are resumed in the scenario "close all, reopen all" (**Figure 9F**). However, already selective opening of only schools bears a strong risk of a second infection wave (**Figure 9D**). **Supplementary Figures S3** and **S4** represent testing of different start times and durations of closures.

In summary, these results have serious consequences for considering the reopening of societies and communities: In this specific simulation, reopening of schools or of all locations after sufficiently long partial closure has no major effect. Following closure of all locations, opening of schools or public spaces can lead to a strong second peak of infection. Here, one has to note that different simulation runs show qualitatively different behavior. Due to the stochastic nature of the model, the infection becomes active in some simulations, but not in others. It is also important to note that in the case of closure of all locations, the number of susceptible individuals stays high, meaning that, even if the virus dies out in the simulation, the simulated population would be at risk of a second wave with just one newly infected person entering the system.

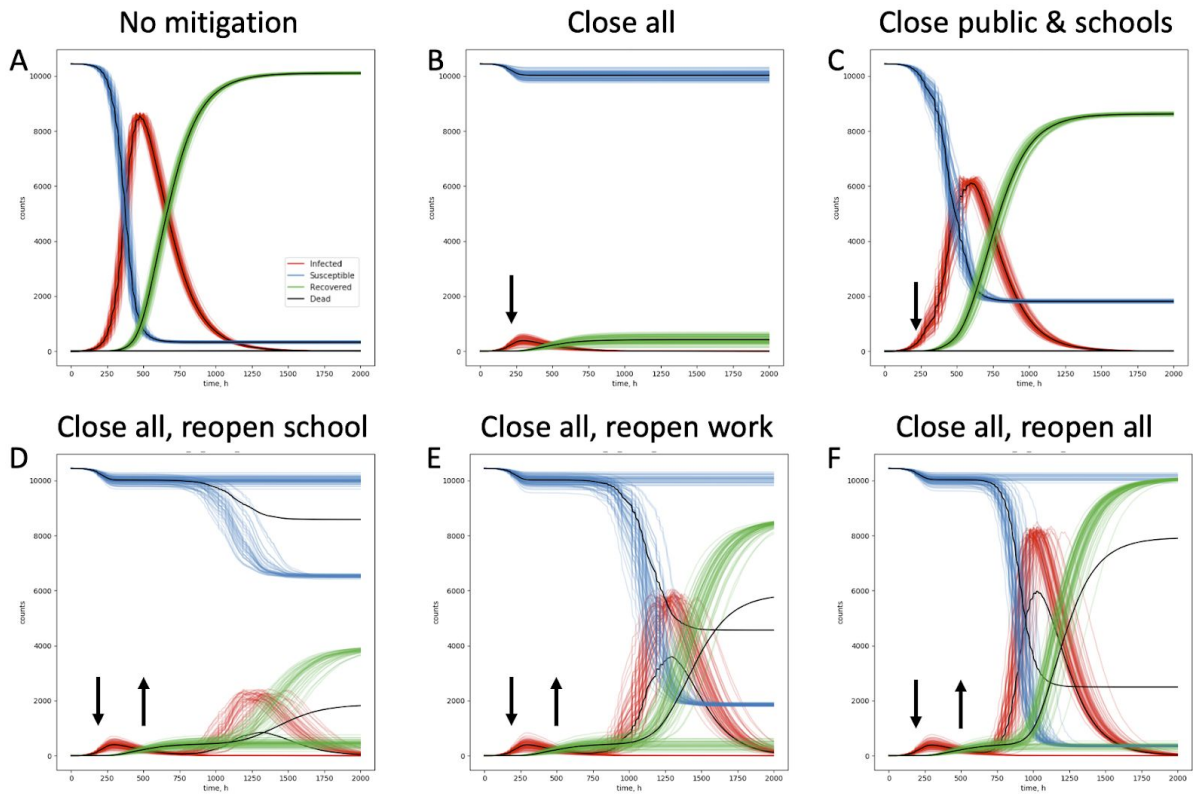


Figure 9. **Comparison of different scenarios of contact restrictions and release.** We compare the effect of no mitigation to scenarios where schools, workplaces (work), public places (public) or all three of them (all) have been selectively closed at 200h and selectively opened at 500h. Closed and opened locations are indicated per panel; downward arrows: closing times, upward arrows: opening times.

The effective reproduction number R_{eff}

The GERDA-1 model can be used to monitor the effective reproduction number R_{eff} in a time-dependent fashion, as an emergent property of the modelled system. The results for the baseline scenario without mitigation and an infection rate $p_{inf} = 0.5$ are shown in **Figure 10**. The value ranges between about 4.8 at the beginning of the infection wave and 0.68 at the end of the simulation period. When we reduce the infection rate to $p_{inf} = 0.3$ mimicking low social distancing, the values range between 3.5 and 0.63. Closure of schools, public and work ("all") leads to faster decline in R_{eff}, however, the final values are slightly higher than without mitigation (about 0.77). The selective reopening of schools after closure of all locations leads to a second increase in R_{eff}, here with values of up to 1.6.

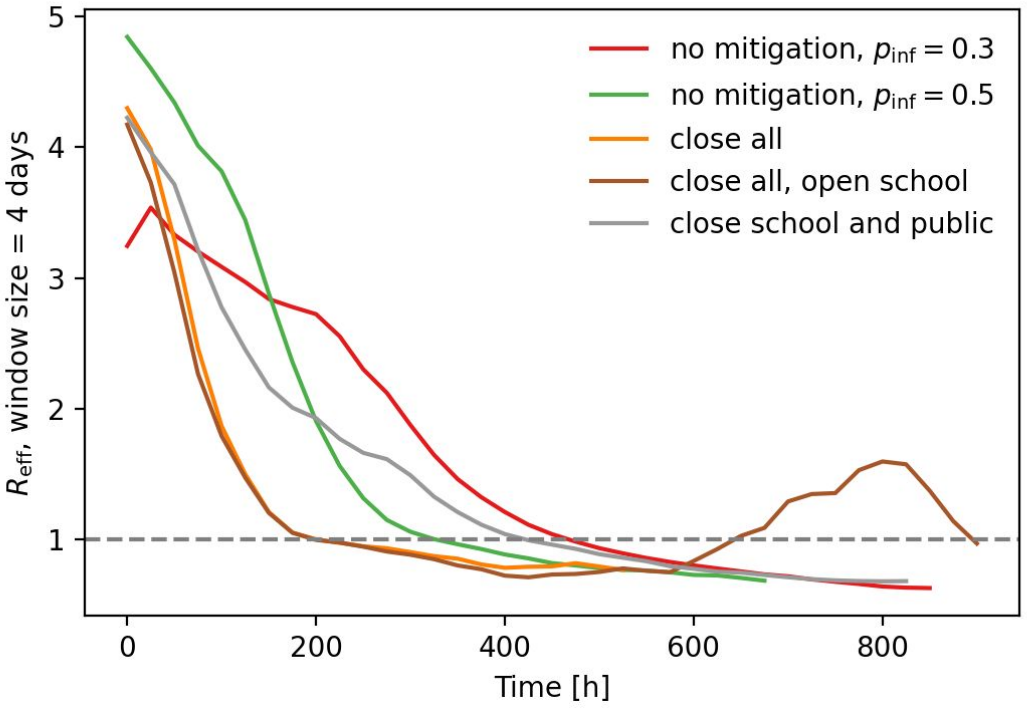


Figure 10. **R_{eff} values for different scenarios.** R_{eff} has been calculated as the average number of individuals who have been infected by one infected individual. Resulting R_{eff} values are average numbers for a sliding window of 4 days (although the total number of newly infected individuals per I-individual was considered). For all scenarios where not specified, the infection rate is 0.5.

Discussion

The presented GERDA-1 model is - to the best of our knowledge - the first geospatial and demography-referenced agent-based model of the Sars-CoV-2-epidemics. It is based on location data and household numbers of a small German municipality but can be readily extended to any community world-wide where data is available. It neither repeats the history in that municipality nor violates the privacy of real individuals, but it relies on realistic assumptions about type and number of locations, distances, number of households and inhabitants as well as their daily routines. It represents a realistic network between the simulated individuals, hence is not entirely random, despite its inherent stochastic nature.

The agent-based model can simulate the stochasticity of virus spreading through a population. Furthermore, its agents are generated based on population data and renders their movements in a realistic environment constructed from real world geodata. Agents have individual behavior based on their age group and encounter other agents by modelled probabilities. This is a significant improvement compared to the classical SIR-models, which are based on ODEs and can, therefore, not simulate individual behavior and randomness nor make use of geolocations. The latter also being an important milestone also for future integration with public health authority recommended smart-phone tracking.

Whilst the model recently published by Kissler et al. [6] is an ODE model fitted with published data and implements social distancing measures, it does not discriminate

between different types of mitigation measures. Our model can discriminate between different types of measures applied to different public places or subgroups of people.

Predicting the outcome of mitigation scenarios. Our model allows us to compare the potential worst case scenario without any protection or mitigation measures (the baseline scenario) to scenarios that represent different protection levels, lockdown measures and reopening strategies.

These simulations allow a set of preliminary conclusions:

1. The effect of lockdown measures importantly **depends on how strict** the measures are implemented and enforced. Lockdown of only public places and schools still leads to a significant infection wave passing through the entire population compared to lockdown of all locations (except of homes and hospitals) where the infection is drastically decreased. Also the effect of a selective opening is dependent on how strict the preceding lockdown has been.
2. If the population still contains a large number of susceptible individuals and sufficient infected individuals, then a selective opening of locations (such as the opening of schools after 500hrs (approx. 20 days) in **Figure 7**) can lead to a **severe second wave** of infection.
3. The infection process is **stochastic** by nature. This means that the effect of reopening cannot be predicted with absolute certainty. Depending on the numbers of S and I as well as on the not fully predictable behavior of the individuals, we can obtain qualitatively different trajectories, meaning that the infection picks up in some cases and decays in other cases.

The role of schools in the infection dynamics. Schools turn out to be a major hub for the progression of infection. Schools are a natural place of contact for the children and teachers. Our simulations (e.g., **Figure 8B,E, Figure 9D**) show that they are also a major location to spread the infection within the municipality and that they contribute essentially to the overall dynamics. For the currently envisaged reopening of schools in many countries the property as a hub and the bimodality of the behavior after reopening means that the outcome of this political decision can be positive (i.e., no major infection wave afterwards) or troublesome (i.e., leading to another major outbreak). However, the presented simulations refer only to a small town (the virtual Gangelt) with only 2 schools and in the current GERDA-1 model version neither kindergartens nor public transport and the effect for larger communities with more inhabitants and travel activities remains to be tested. To the best of our knowledge, the GERDA-1 model is the first model able to predict the bimodal behavior of SARS-Cov-2 infection dynamics.

Exposure of specific groups of the population. Health care workers and medical professionals are at the forefront fighting COVID-19. Our model also reveals the importance of specifically protecting them since a lack of personal protective gear leads to increased numbers of infected and deceased medical staff, but also to an elevated spreading of the disease in the

whole population. In addition, it can be expected that this trend is enhanced upon including detailed hospital capacity metrics (staff, beds, ICU beds) in the model.

Potential limitations of the model. Here, we have integrated real demographic *data* reflecting the number of households, composition of households (singles, families, couples, shared apartments), age distribution. We also include relevant data about daily activities of individuals such as being at home, visiting schools or workplaces or engaging in social activities. Specifically, we also consider individuals working in healthcare (first-responders, nurses and doctors). This allows to test much more specific scenarios of mitigation or the effect of relaxation measures than do classic epidemiological models, which describe only the different epidemic groups without reference to their real life occupations or geopatterns. Note that we modeled the outbreak in only one small-town municipality, the dynamics might differ if more connected regions are included (metropolitan areas, suburbs etc).

How trustworthy are the parameter values? We have derived all relevant transition probabilities from currently available information for the propagation of COVID-19 (see Supplementary Table S1, [4]). However, this information is neither complete (w.r.t. the transitions that we have in the model) nor is it stabilized. For example, diagnostic capacity in Germany is increasing while we simulate and write this manuscript, hence, the respective parameters need to be adjusted during the pandemic and for different geolocations. The infection probability (whether an S-agent gets infected by an I-agent when meeting at the same location) by itself is hard to measure directly (as statistics over infection events are lacking). Thus, the respective values is best educated guess and we will test the effect of this parameter further with sensitivity analysis. On the other hand, the number of locations has been taken from current databases (OpenStreetMap) so can be considered rather robust. Composition of households is from the German census data, thus trustworthy, however potentially not exactly representing the conditions for every region.

For comparison, the latest numbers reported from Gangelt of April 30 2020 (status at 15:00) [https://www.kreis-heinsberg.de/aktuelles/aktuelles/?pid=5149] are 478 diagnosed individuals, 446 recovered individuals and 4 deceased individuals since February 25 2020.
The current version of GERDA-1 entails only a *limited* set of opportunities for *daily schedules* and *locations* for each agent. These were schedules we considered most representative. In forthcoming versions of the model we will allow for more diverse and adapted schedules. First, we will include nursery/senior homes to account for a location with potentially drastic consequences during the pandemic. Then we will also take into account daily occupations such as kindergartens, differentiate between elementary schools and high schools, but also enclose universities and other educational instances. We will also allow for more diversity of workplaces (SME vs large-industry), e.g. in public transport, public service, or shops compared to workplaces with more limited/repetitive contacts such as offices or factories. These data will be drawn from the respective statistics.

Generalization and potential extensions of the model

The model concept allows for a multitude of extensions. Our goal will be to extend the georeference of the model to enable the EMEA/EU and subsequently global community to run simulations via a web portal. We hope Alphabet Inc. will open up for access to Google Maps. In future versions of the model, we consider to include further specialized groups, such as system-relevant workers, workers with direct client contact or shift workers. This will provide the basis for simulating diverse mitigation measures in very different populations. Incorporating further locations, especially to represent different public spaces, such as but not restricted to super markets, public transport will lead to simulation of more diverse municipalities. In consequence, expansion of the model to several small municipalities or bigger cities will be feasible.

A *stepwise extension* of the model will comprise (pursued in parallel collaborating developer-teams) (i) other cities and municipalities in Germany, (ii) a combination of several cities in Germany, (iii) whole Germany, and as soon as possible (iv) Europe and worldwide, and (v) adaptation and improvement of individual behavior. However, it is clear that each extension will require data representing locations and inhabitants of the considered areas and that each extension will also demand more development time, computing power, and testing capacity.

While our model and efforts to implement it, has been very much motivated by the current corona crisis, the presented concept and the planned extensions are also suited to analyze the distribution of other infectious diseases in an unprecedented precision, since we include demographic and geographic data that is as precise as current databases allow and that also determines spreading of, for example, the common flu or other coronaviruses and pathogens.

Material & Methods

The ABM and the simulation framework were designed in an object-oriented manner, using the programming language **Python** version 3 [7] and the packages **Numpy** [8], **Pandas** [9], **Geopandas** [10], **osmnx** [11]. The package **matplotlib** [12] was used for visualisation. Despite being streamlined for the parallel execution of numerous replicates; the developed tool was designed for usage on customary machines.

Data Sources

John Hopkins Corona resource:
https://coronavirus.jhu.edu/map.html

Ourworldindata
https://ourworldindata.org/coronavirus#all-charts-preview

IHME and CDC, ECDC.

Robert Koch-Institut (RKI)

https://www.rki.de

1. Current Stage/Situation report of Robert-Koch-Institute (RKI) about COVID-19:
   1.1. COVID-19 fatalities reported to RKI sorted according to age and sex
   1.2. Reported COVID-19 cases per 100.000 inhabitants in Germany sorted according to age groups and sex
   1.3. Case numbers
   1.4. Intensive care numbers / see Divi Intensive Register

2. SARS-CoV-2 Characteristics for Coronavirus-Disease-2019 (COVID-19) (in German) https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Steckbrief.html
   2.1. - Data from the characteristics refer to different studies

3. Modellierung von Beispielszenarien der SARS-CoV-2-Epidemie 2020 in Deutschland https://edoc.rki.de/handle/176904/6547.2

Divi-Intensivregister

https://www.divi.de

1. Tagesreport DIVI Intensivregister
2. Case numbers of reported intensive care patients (to estimate transition probabilities)

Statistisches Bundesamt

Census, Distribution of household types, size and corresponding age distributions

1. https://ergebnisse.zensus2011.de/
2. https://service.destatis.de/

OpenStreetMap

Geospatial data for initialization of locations openstreetmap

https://www.openstreetmap.de/

Code Availability

The code will be available shortly at

https://ford.biologie.hu-berlin.de/tbphu/corona_abm

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG: Cluster of Excellence MATH+, TRR 175) and by the German Ministry of Education and Research (BMBF,

Liver Systems Medicine (LiSyM) network grant) and by the People Programme (Marie Skłodowska-Curie Actions) of the European Union's Horizon 2020 Programme under REA grant agreement no. 813979 ('Secreters'). Escalera-Fanjul X is supported with a postdoctoral grant from CONACYT (CVU 420248). We would like to also thank Luke Sherman for his help in proofreading this manuscript.

References

[1] Kermack, W. O.; McKendrick, A. G. (1927). "A Contribution to the Mathematical Theory of Epidemics". *Proceedings of the Royal Society A*. **115** (772): 700–721.

[2] B. F. Maier and D. Brockmann. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China Science (2020) doi:10.1126/science.abb4557

[3] Stonedahl, F. and Wilensky, U. (2008). NetLogo Virus on a Network model. http://ccl.northwestern.edu/netlogo/models/VirusonaNetwork. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

[4] He, X., Lau, E.H.Y., Wu, P. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0869-5

[5] Handel A, Longini IM Jr, Antia R. What is the best control strategy for multiple infectious disease outbreaks?. *Proc Biol Sci*. 2007;274(1611):833-837. doi:10.1098/rspb.2006.0015 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2093965/gpl

[6] Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Published Online14 Apr 2020 doi: 10.1126/science.abb5793

[7] Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

[8] Oliphant, T. E. (2006). *A guide to NumPy* (Vol. 1). Trelgol Publishing USA.

[9] McKinney, W., & others. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51–56).

[10] geopandas: https://geopandas.org/

[11] Boeing, G. 2017. "OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks." Computers, Environment and Urban Systems 65, 126-139. doi:10.1016/j.compenvurbsys.2017.05.004

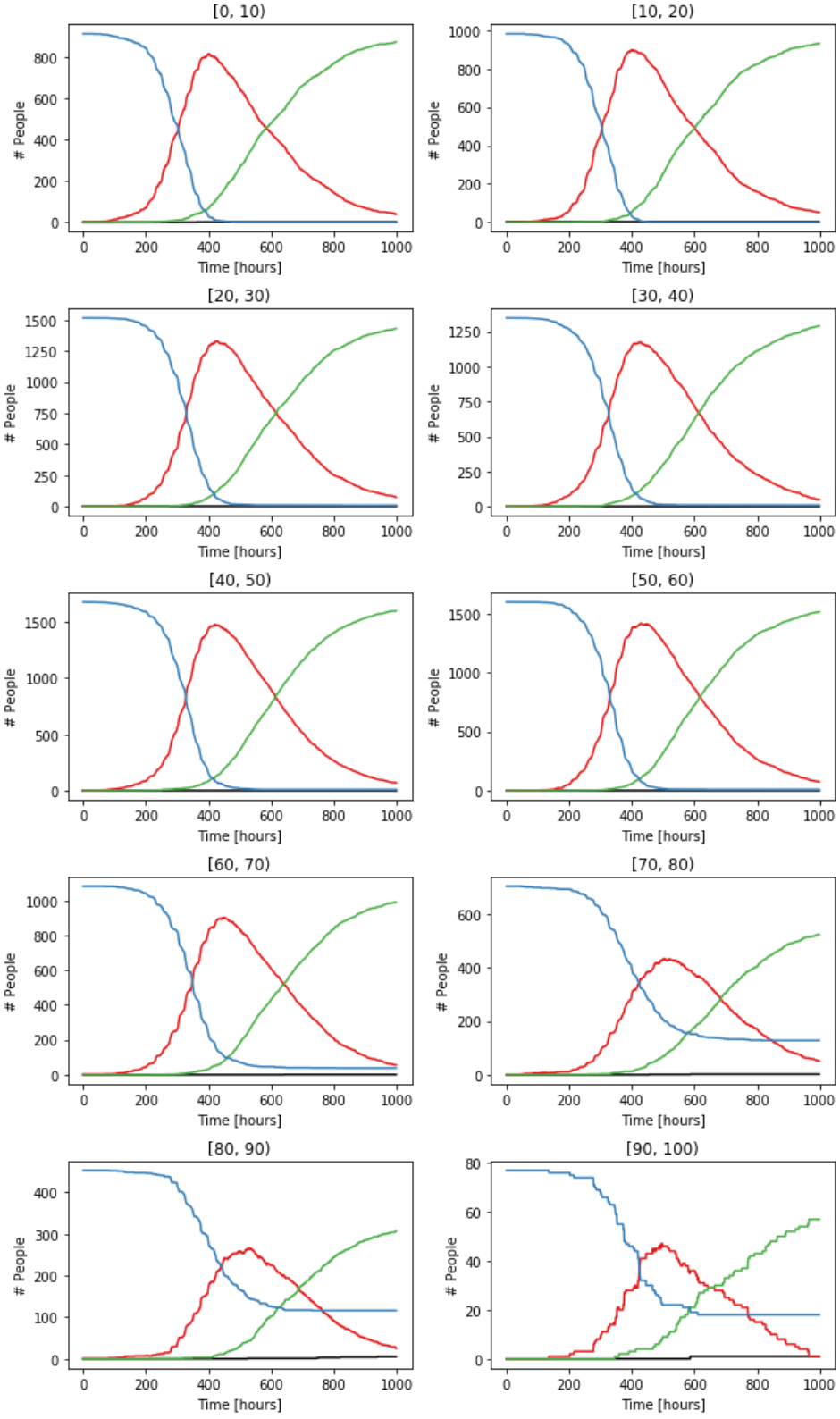
[12] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95.

Supplementary Information

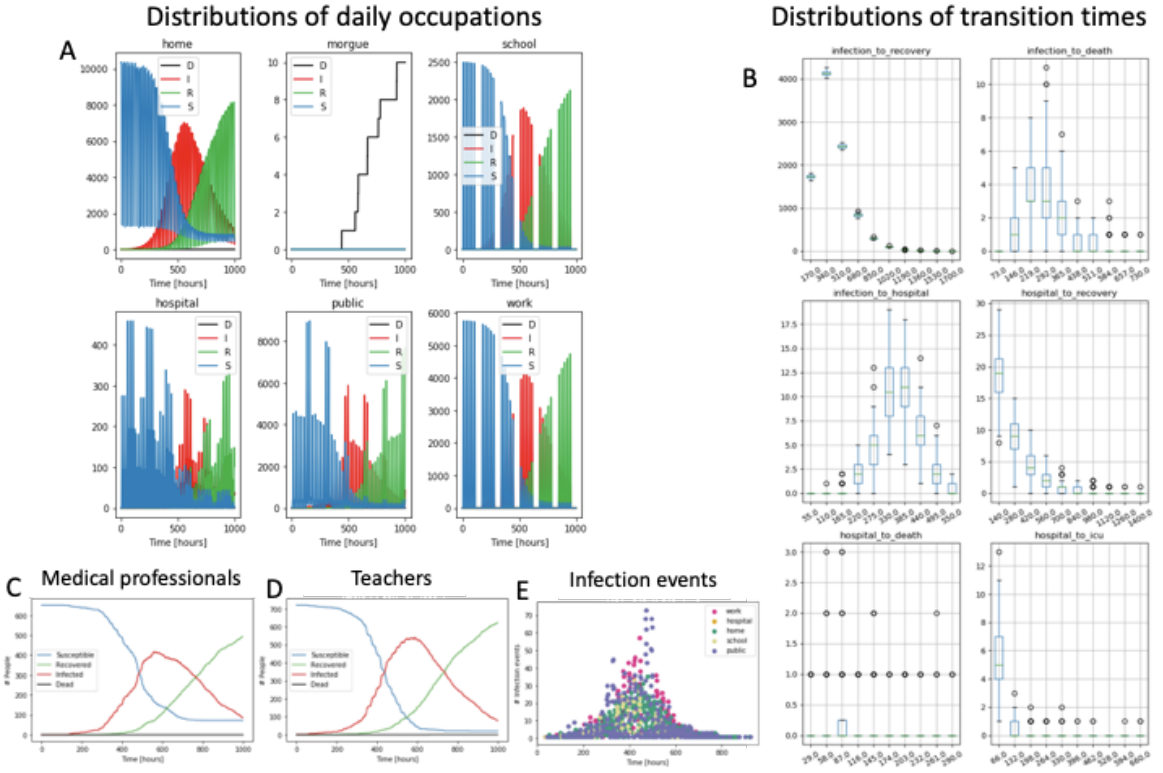
Supplementary Table S1: Quantitative Information used for state changes. For the state changes in the left column we derived the age-dependent probabilities per hour for the transition to occur from the following data: average time in a state until the transition (given in days) and percentage of affected people undergoing the transition.

State Change	Average Time to transition [d]	Percentage of individuals affected	Source
Infected → Recovered	10	85 %	RKI
Infected → Hospital	5.5	15 %	RKI
Infected → Death	10	Age Dependent Death Rate	RKI
Hospital → Recovered	14	78 %	RKI
Hospital → ICU	14	22 %	DIVI-Register
Hospital → Death	10	Age Dependent Death Rate	RKI
ICU → Hospital	7	70 %	DIVI -Register
ICU → Death	7	30 %	DIVI-Register
Susceptible → Infected	Probability distribution taken from [1]		
Infected → Diagnosed	15	15 %	RKI

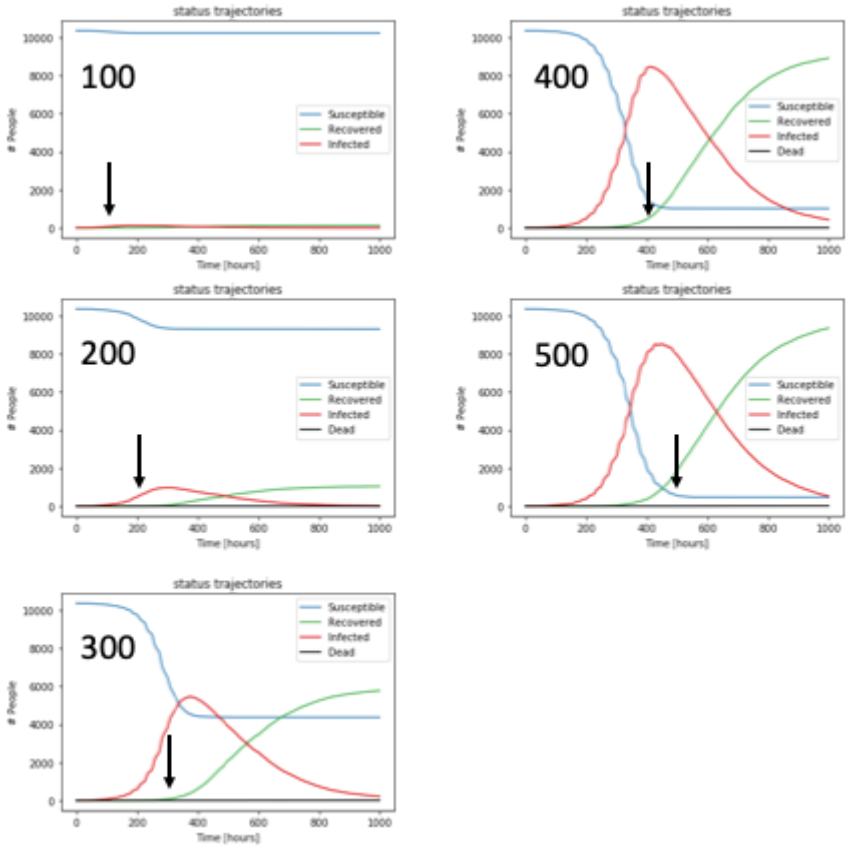
Supplementary Figure S1: Age-dependent trajectories for S, I, R, and D. Curves show results for a single simulation for infection rate 0.5. Age ranges are given on top of each panel.



Supplementary Figure S2: Quantitative characterization of dynamics for infection rate 0.3 without mitigation measures. **A**) Distribution of daily occupations. **B**) Distribution of transition times. **C**) Time courses S, I, R, and D for medical professionals, **C**) Time courses S, I, R, and D for teachers, **E**) Number of infection events at given times per location type. Panel **B** shows the statistics for 100 simulation runs, Panels **A**, **C**, **D**, and **E** show results for a single representative simulation run.



Supplementary Figure S3: Variation of the onset time for complete shutdown (Close_all modus). Shown is the scenario with closure of schools, work and public at the times given in the panel (i.e., 100h up to 500h).



Supplementary Figure S4: Variation of start times of closure of all locations and durations of complete shutdown followed by school reopening. Parameter value $P_{inf} = 0.5$.