scientific reports

OPEN

Inferring the effect of interventions on COVID-19 transmission networks

Simon Syga[1], Diana David-Rus[2], Yannik Schälte[3,4], Haralampos Hatzikirou[5] & Andreas Deutsch[1]✉

Countries around the world implement nonpharmaceutical interventions (NPIs) to mitigate the spread of COVID-19. Design of efficient NPIs requires identification of the structure of the disease transmission network. We here identify the key parameters of the COVID-19 transmission network for time periods before, during, and after the application of strict NPIs for the first wave of COVID-19 infections in Germany combining Bayesian parameter inference with an agent-based epidemiological model. We assume a Watts–Strogatz small-world network which allows to distinguish contacts within clustered cliques and unclustered, random contacts in the population, which have been shown to be crucial in sustaining the epidemic. In contrast to other works, which use coarse-grained network structures from anonymized data, like cell phone data, we consider the contacts of individual agents explicitly. We show that NPIs drastically reduced random contacts in the transmission network, increased network clustering, and resulted in a previously unappreciated transition from an exponential to a constant regime of new cases. In this regime, the disease spreads like a wave with a finite wave speed that depends on the number of contacts in a nonlinear fashion, which we can predict by mean field theory.

The SARS-CoV-2 pandemic has dramatic consequences at a global scale. Until herd immunity has been reached through vaccination, countries rely on non-pharmaceutical interventions (NPIs) of varying severity, like canceling big events, closing schools, and shutting down businesses to reduce virus transmission. An important goal of NPI design is to prevent those contacts in the population which contribute the most to disease spread while allowing less dangerous contacts. Since the start of the pandemic, mathematical models have been playing an important role in guiding policy makers by developing scenarios for the number of cases, hospitalizations and deaths[1,2], and estimating the effects of NPIs on the spreading dynamics[3–6]. They are also used to estimate the herd immunity threshold[7]. However, the effect of NPIs on the transmission network of the disease and the corresponding changes in the spreading dynamics are still not fully understood.

In several countries, following the implementations of various NPIs, the curves of cumulative cases left the exponential regime and entered an approximately linear one, corresponding to a constant number of daily new infections, and an effective reproduction number around one, see Fig. 1a–c. This is remarkable, because these countries are heterogeneous regarding their demographics, economic situation and the implemented NPIs. Epidemiological models that assume a fully connected transmission network, like compartment models based on ordinary differential equations (ODEs), predict this behavior only for a particularly fine-tuned set of parameters, which contrasts with the robustness of the decline in the reproduction number observed in reality[8,9], see Fig. 1d–f. This is still true for detailed compartment models that incorporate the effect of test–trace–isolate (TTI) efforts, asymptomatic spreaders, or age-dependent spreading based on contact matrices[4,10] and for agent-based models that rely on coarse-grained contact networks, for example created from anonymized cell phone data, which assume a well-mixed situation on a mesoscopic scale of hundreds of agents[5,11,12]. Komarova et al.[13] discussed the power law behavior of the dynamics of COVID-19 spread in the context of a metapopulation model, where the population is divided into connected patches assuming a fully connected transmission network within individual patches. The power law dynamics during the hard lockdown in Chinese provinces could be explained by

[1]Center for Information Services and High Performance Computing, Technische Universität Dresden, Nöthnitzer Straße 46, 01062 Dresden, Germany. [2]Bavarian Health and Food Safety State Authority (LGL), Veterinärstraße 2, 85764 Oberschleißheim, Germany. [3]Institute of Computational Biology, Helmholtz Zentrum München – German Research Center for Environmental Health, 85764 Neuherberg, Germany. [4]Center for Mathematics, Technische Universität München, 85748 Garching, Germany. [5]Mathematics Department, Khalifa University, P.O. Box 127788, Abu Dhabi, UAE. ✉email: andreas.deutsch@tu-dresden.de

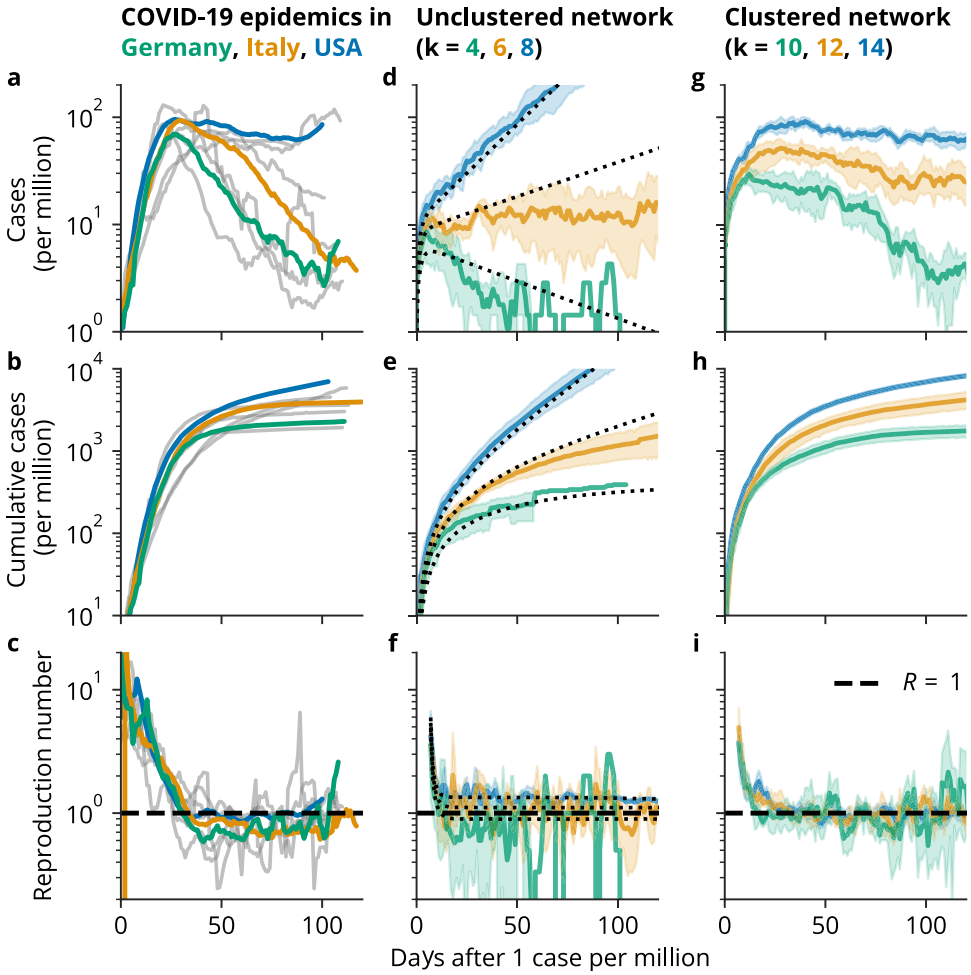

Figure 1. Dynamics of COVID-19 epidemics compared to model dynamics in hypothetical clustered and unclustered networks. (**a–c**) Seven-day rolling average of new cases per million, cumulative cases per million and reproduction number in Germany, Italy and the US, aligned to the time point when the 7-day rolling average reached one per million. After the implementation of NPIs, case numbers decrease slowly, corresponding to effective reproduction numbers of just below one. As an extreme example, in the US, case counts stayed almost constant for approximately 2 months, before increasing again. In contrast, in Germany, cases decreased by about 90% in 2 months, similar to other Western European countries (gray lines). Case counts in Italy, the epicenter of the first wave, are between those of the US and Germany. (**d–f**) Disease dynamics in a random network. The number of new cases and cumulative cases changes exponentially over time, strongly depending on the number of contacts and the infection probability. The reproduction number is equal to one only for a fine-tuned set of parameters. Black dotted lines correspond to predictions of the respective differential equation approximation. (**g–i**) Disease dynamics in a strongly clustered small-world network ($p \approx 0$). After an initial exponential increase in cases, the number of new cases is almost constant over time, corresponding to a linear increase in cumulative cases and reproduction numbers around one. This behavior is robust against changes in the total number of contacts k and the infection probability p_I. (**d–i**) show mean and standard deviation of 5 independent simulations per parameter set on Watts–Strogatz networks with $n = 10^5$ nodes. The data in (**a–c**) is provided by Johns Hopkins University[28].

a compartment model that assumes that the susceptible population was quarantined on a time scale comparable to the infectious period of the disease, so that the epidemic comes to a halt quickly[14]. However, this assumption is hard to justify for the situation in Western countries like Germany, that did not implement a full lockdown but enacted contact restrictions and closure of nonessential businesses so that people were still allowed to leave their homes and meet in small groups.

In contrast to models that implicitly assume fully connected transmission networks, network-based epidemiological models allow to consider the effects of heterogeneity with respect to the type and frequency of contacts in the population, i.e. how often people meet and whom, by representing all agents as nodes of a network and the contacts in the population by links between these nodes[15,16]. The heterogeneous topology of real social networks is reflected by a small average path length between any two nodes (small world property), a high clustering in the network (the probability of two nodes being connected is much larger if they have a neighbor in common)

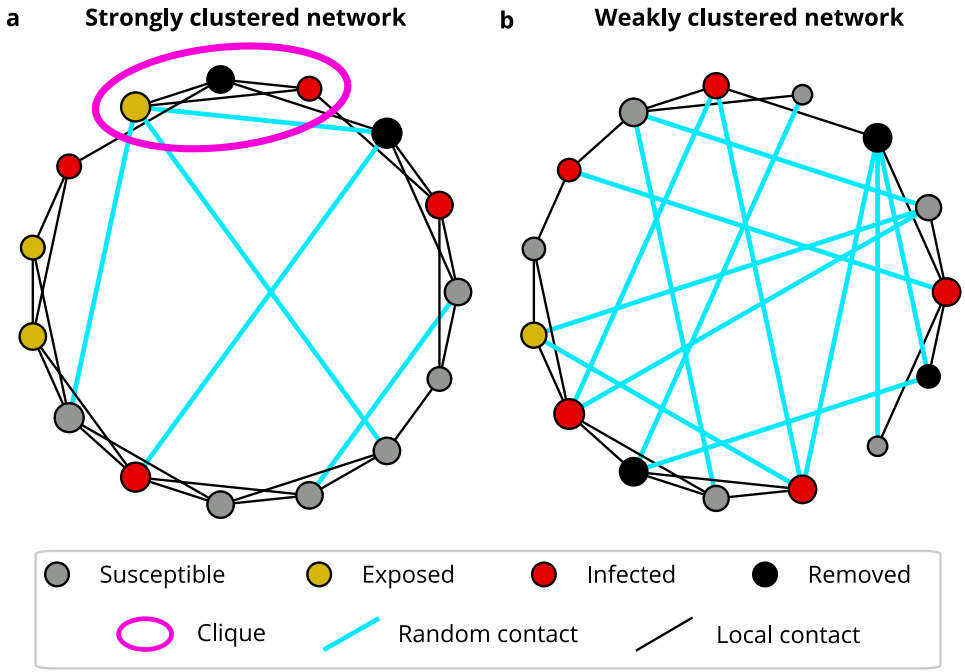

Figure 2. Watts–Strogatz small world network. Agents are placed in a ring-like topology and linked to their k nearest neighbors (black lines). Next, every link is rewired randomly with a small probability p (cyan lines). Every agent has one of four states: susceptible (gray), exposed (gold), infectious (red), or removed (black). Infectious agents spread the disease to connected susceptible agents with a probability p_I in each time step. The size of network nodes is proportional to the node degree. (**a**) Strongly clustered network. Almost all contacts are restricted to neighbors ($p = 0.1$). (**b**) Weakly clustered network with a large fraction of random contacts ($p = 0.5$). Other parameters are equal in (**a**) and (**b**): $n = 15$ agents, $k = 4$ average contacts.

and by a power-law distribution (scale-free property) of the node degree[17,18]. The structure of the transmission network should be considered in mathematical models, because it can change the observed spreading dynamics[19]. For example, the spread of diseases is strengthened on scale-free networks so that the epidemic threshold is reduced[20–22]. Thurner et al.[23] suggested that the linear regime of cumulative cases is a consequence of small-world transmission networks with high clustering, see Fig. 1g–i. A similar system was studied in a theoretical work on epidemic spread with two levels of mixing[24]. There, the authors found that transmission within interconnected cliques leads to an increase in the effective reproduction number proportional to the number of infected people in the cliques. This in turn means that a very small number of links between cliques is enough to enable a large outbreak, which corresponds to the exponential regime. A model based on a spatial transmission network with a variable interaction range also showed power-law growth of new cases[25]. A related model could explain the disease dynamics during the SARS outbreak in Hong Kong in 2003[26].

However, while previous works on network-based models for COVID-19 focus on qualitative aspects of certain network ensembles, like power-law growth[19,23,25], we here aim to explicitly infer the time-dependent transmission network for COVID-19 in Germany. We argue that during the period of severe NPIs, like contact restrictions, the most important feature of real transmission networks is their strong clustering. This means that because public places and events are closed, we expect that people focus their contacts on a single group (clique), where almost each member of this clique is contact with each other. Typical examples of such cliques include households or teams at work. On the other hand, we will neglect the scale-free property of social networks, because it requires that there are a few people with a very large number of contacts, for example at events, schools, large private gatherings etc., which are the targets of most NPIs. We combine Bayesian parameter inference[27] with an epidemiological model based on the Watts–Strogatz small-world network[17] that allows to interpolate between unclustered and highly clustered transmission networks by varying the fraction of random contacts in the population, see Fig. 2, to infer the topology of the transmission network in Germany during three time periods in 2020: February 26 until March 15, before serious NPIs were imposed, March 16 until June 6, when strict contact restrictions were in place and nonessential businesses were closed, and June 7 until September 15, when most NPIs were lifted. Random contacts often span a large distance in the transmission network and connect different cliques. They include, for example, contacts in public transport, bars and restaurants, but also contacts with relatives that live far away. Furthermore, given the nature of random contacts, the probability of superspreading events increases when there is a high density of such random contacts, as they enable the disease to spread to fully susceptible cliques.

After inferring the model parameters, we perform a parameter scan to identify the transition points between the linear and the exponential regime of the dynamics. This allows us to associate the three time periods with

regions in the phase diagram. Next, we derive a mean-field analysis of a simplified model to gain a deeper understanding of the dynamics in the linear regime. We end with a critical discussion of our results.

Results

Bayesian parameter inference. We aimed to infer the induced changes in the topology of the COVID-19 transmission network by Bayesian parameter inference. We expected that NPIs lead to a change in the behavior of people and thus in the topology of the corresponding transmission network. We assumed that the transmission network can be described by the Watts–Strogatz network[17] that can interpolate between a weakly and a strongly clustered small-world network, see Fig. 2. Crucially, in this framework, we could distinguish local, clustered contacts within cliques, like households, nursing homes, businesses, etc. and random contacts outside of these clusters, corresponding to encounters in public transport, with business partners, friends and family that live far away and similar. During the construction of the Watts–Strogatz network, n nodes are placed in a ring topology and connected to their $k = 2, 4, 6, \ldots$ nearest neighbors (local contacts, black lines in Fig. 2). After that, each link is rewired with a probability p to another random node (random contacts, cyan lines in Fig. 2). We used the SEIR (susceptible-exposed-infectious-removed) epidemiological model for the COVID-19 disease dynamics. Agents in the SEIR model were represented by network nodes such that infectious agents could spread the disease with probability p_I to the susceptible agents that they are connected to in discrete time steps of single days ("Methods" section, Fig. 2, Supplementary Fig. 1).

We inferred the model parameters p, k, p_I for the time periods before and after the first NPIs were implemented in Germany, and after most NPIs were lifted again using an approximate Bayesian computation with sequential Monte Carlo (ABC-SMC) algorithm ("Methods" section). See Table 1 for an overview of the model parameters. We did not infer the mean time periods agents spend in the exposed and infectious states, because they were reported in the literature as 3d ± 1d (mean ± standard deviation) from exposure to becoming infectious and 10d ± 3d of being infectious[29]. We kept the number of agents fixed at $n = 3 \times 10^5$, which we regard as a representative sample of the whole population. Note that we do not explicitly account for the quarantine of infectious agents, which could be done by a time-dependent removal probability, for example. However, we do not do this, because the infection and removal probabilities cannot be determined independently[3] and there is no reliable data on the effectiveness of TTI measures to further specify the removal probability.

For the time periods before June 6, for which a large number of infections was undetected, we also inferred the initial numbers of exposed and infectious individuals $n_E(0), n_I(0)$.

Based on Google mobility data and previous work on the inference of change points in the spread of COVID-19[3,30] we assumed the critical time point for the effect of NPIs in Germany to be March 15, as from March 16, NPIs were synchronized in German states, and schools and nonessential businesses were closed.

We intentionally chose broad, uninformative priors for all parameters, such that we could compare the obtained posterior distributions with other data sources as a sanity-check of our approach. To account for the weekday-dependent reporting delay, we used a seven-day rolling average of new case reports provided by Johns Hopkins University[28], see Fig. 3a. Our parameter inference scheme is based on a minimization of the difference between this average and the number of agents becoming infectious in the model on the corresponding day. For the initial phase, we base our inference on the absolute number of infections, while we used the relative number for the following time periods (Supplementary Information).

First, we inferred the parameters for the time from February 26 to March 15, since daily new cases increased rapidly after February 26, while there were almost no cases in the week before. Our analysis of this period revealed an almost random transmission network, with a median fraction of random contacts of $p = 0.48$ (with 95% credibility interval, CI [0.23, 0.94]), a large number of contacts $k = 26$ (CI [22, 32]) and a high infection probability of $p_I = 0.035$ (CI [0.025, 0.061]). The inference of this unexpectedly high number of contacts could be the result of a scale-free degree distribution before NPIs were imposed (Discussion). For the initial condition we estimated that 33 (CI [4, 46]) people were exposed and 81 (CI [32, 118]) people were infectious on February 26.

To assess the changes of the transmission network induced by the NPIs in Germany, we next considered the time period following March 16. During that time frame, the NPIs were changed several times, however, the contact restrictions, which we regard as the most crucial intervention, were only lifted on June 6, which is why we chose this date as the endpoint of the time interval. As the total number of cases was computationally intractable, we here used the relative number of new cases as input for this time period (Supplementary Information).

Our Bayesian parameter inference revealed that the NPIs reduced the number of contacts in the transmission network considerably to $k = 6$ (CI [4, 10]). They also reduced the infection probability of these contacts to $p_I = 0.02$ (CI [0.010, 0.039]), which matches well with an estimation based on the individual-level secondary attack rate in the household of 17%[31]. Crucially, the fraction of random contacts decreased to $p = 7 \times 10^{-5}$ (CI $[10^{-7}, 0.12]$), stopping the exponential growth. Additionally, we estimated the number of exposed people on March 16 to be $n_E(\text{March16}) = 78$ (CI [31, 147]) per million and the number of infectious people to be $n_I(\text{March16}) = 452$ (CI [301, 656]) per million. The fact that during the week before March 16 there were only 57 infections per million detected in Germany is a hint that a large fraction of infections went unnoticed at the time, which agrees with other reports[32,33].

We also inferred the model parameters for the time period following June 6 when contact restrictions were lifted. To this end, we used the final time point of simulation instances from the previous period as initial conditions. For the number of contacts, we obtained a median $k = 12$ (CI [6, 20]) that matches well with reports of the average number of daily contacts in Europe of 13.4[34]. The median fraction of random contacts was estimated as $p = 0.03$ (CI [0.001, 0.6]), which means it was notably smaller than before NPIs had been implemented, but larger than in the time period of strict NPIs. Interestingly, we found that the infection probability was as low in this time period at $p_I = 0.02$ (CI [0.01, 0.04]) as during the lockdown, which could be the result of a seasonal

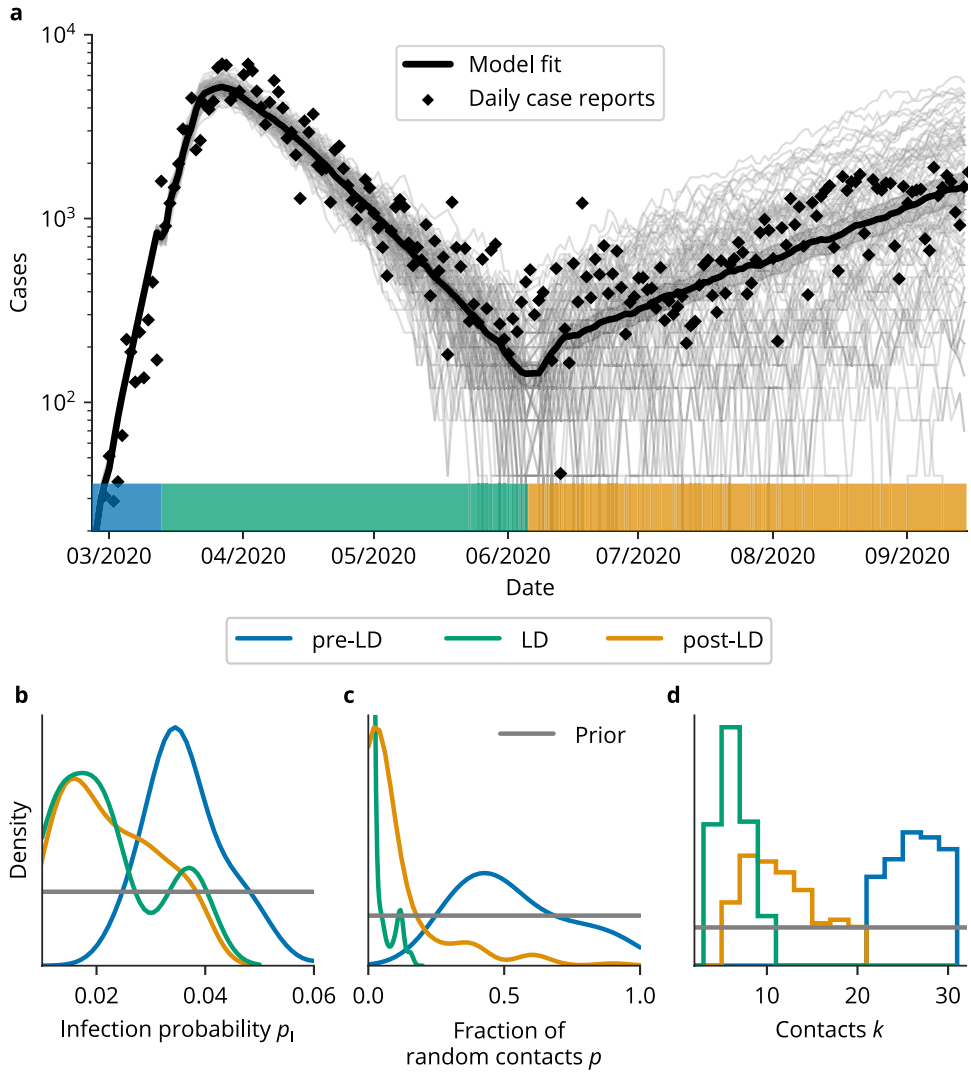

Figure 3. Inference of key epidemiological parameters before (pre-LD, blue), during (LD, green) and after (post-LD, yellow) the lockdown in Germany. (**a**) Daily case reports (black diamonds), single model instances (gray lines) and their mean (black line, error band corresponds to 95% error of the mean). Colors indicate the respective time periods, where the color is the same as that of the corresponding posterior parameter distributions in (**b–d**). Model parameters were chosen as the median of the respective posterior distribution. (**b–d**) Kernel density estimates of the posterior parameter distributions. The uniform prior distribution is shown in gray. (**b**) The infection probability p_I decreased from $p_I = 0.035$ (CI [0.025, 0.061]) pre-lockdown to $p_I = 0.02$ (CI [0.01, 0.04]) during the lockdown. After most restrictions were lifted, the infection probability remained almost unchanged at $p_I = 0.02$ (CI [0.01, 0.04]). (**c**) The fraction of random contacts in the transmission network p decreased strongly from $p = 0.48$ (CI [0.23, 0.94]) to $p = 8 \times 10^{-5}$ (CI $[10^{-7}, 0.12]$) when restrictions were put in place and increased to $p = 0.03$ (CI [0.001, 0.60]) when they were lifted. (**d**) The total number of contacts k decreased strongly from $k = 26$ (CI [22, 32]) to $k = 6$ (CI [4, 10]) during lockdown before increasing again to $k = 12$ (CI [6, 20]) after restrictions were lifted.

effect and of people spending more time outside, which hinders the spread of airborne diseases such as COVID-19. We inferred rather broad parameter posterior distributions for this period, due to the generally low number of infections and large localized outbreaks, leading to a large variation in daily case counts. This is also reflected in a large variability between single model instances for this time period, see Fig. 3a (gray lines). For the time periods without strict NPIs (February 26 to March 15 and June 6 to September 15), for which we inferred a large fraction of random contacts, the infection probability p_I and the number of contacts k were notably correlated, leading to broader posterior distributions. Fig. 3 shows the model fit compared to the daily case counts and the corresponding prior and posterior parameter distributions of the network parameters and the infection probability for the three time periods.

Disease dynamics in the clustered network. To determine the transitions between the linear and exponential regimes of the dynamics, we performed a parameter scan varying the network parameters p and k,

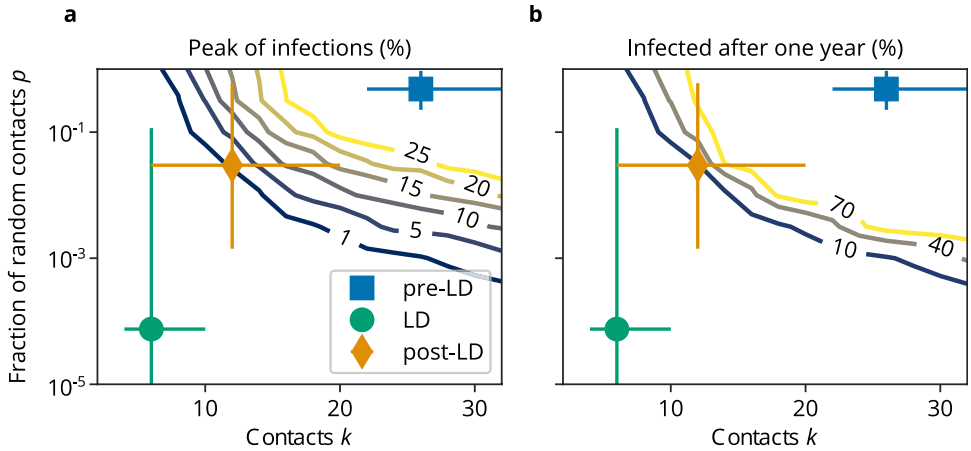

Figure 4. Model dynamics in dependence on the total number of contacts k and the fraction of random contacts p. (**a**) Number of simultaneously infected people (peak of infections) in percent of population. The wave peak can be massively mitigated by decreasing the fraction of random contacts, even while keeping the total number of contacts constant. (**b**) Cumulative number of infections after one year, in percent of the population. Similar to the peak of infections, the cumulative number of infections can be limited by reducing the fraction of random contacts. Blue square (pre-LD), green point (LD) and yellow diamond (post-LD) correspond to the median parameters and 95% CI obtained from Bayesian parameter inference for the time periods 26/02–15/03, 16/03–05/06 and 06/06–15/09, respectively, and are shown as a reference. The NPIs after March 15 prevented an exponential spread of the disease, but lifting them led to another exponential increase. Shown are contour lines of the mean of 20 independent model realizations of each parameter combination (p, k), while the other parameters were fixed at $n = 10^5, p_I = 0.02$.

while keeping the disease-specific parameters and the system size fixed at $n = 10^5, p_I = 0.02$. Thereby, reducing k corresponds to reducing the total number of contacts (local and random), while reducing p does not change the number of contacts, but restricts them to locally clustered agents (cliques). We recorded the peak number of simultaneously infected people, because a central goal of NPIs is to prevent the overload of the health system, and the total number of infected people after 100 days as a measure for the total damage to public health, see Fig. 4.

Importantly, the number of infections could be reduced massively by only decreasing the fraction of random contacts in the population, while keeping the total number of contacts constant. As we have inferred the parameters of the transmission network for different time periods, we could associate them with regions in our parameter space. If no NPIs had been implemented and people would not have changed their behavior, more than 40% of people could have been infected simultaneously and almost everybody would have been infected after one year (Fig. 4, blue square). The peak of infections was reduced to 0.0002% by the interventions (Fig. 4a, green point). Lifting the NPIs moved the system back into the exponential regime, with a projected peak of infections of 1.3% and almost 10% of the population to be infected within one year (Fig. 4, yellow diamond). Reducing the total number of contacts can in principle push the system below the epidemiological threshold leading to extinction of the disease (see Figs. 1, 4), however the effect is weaker in the regime far away from the threshold (see Figs. 1, 4, $k = 10, 12, 14$). On the other hand, in the strongly clustered regime $p \approx 0$, increasing the fraction of random contacts has a dramatic effect: both the peak value of infected agents and the total number of infected agents increase in a non-linear manner. Preventing most random contacts in the network ($p \to 0$) hinders the spread of the disease, so that the effective reproduction number fluctuates around 1, and the cumulative number of infections increases linearly with time as observed in many countries after the first NPIs were imposed.

Wave speed of infections in the linear regime. We were especially interested in the disease dynamics in the regime $p \to 0$, as this is where traditional epidemiological models that assume a random transmission network break down. In the case of only local contacts in the network, the disease spreads like a wave originating from the initially infectious agent. This wave-like disease spread was also reported in real networks, such as the air traffic network[35]. To calculate the speed of the infection wave, we used a mean-field approximation of an SIR-like agent-based model operating on the Watts–Strogatz network (Supplementary Information). We scaled the mean-field equations to continuous time t and space x (where the distance Δx is measured as the number of links along the ring of nodes), and approximated the dynamics by a set of partial differential equations for the probability densities of susceptible $\sigma(x, t)$, infectious $\iota(x, t)$ and removed $\rho(x, t)$ agents (Supplementary Information). In particular, we obtained the following equation for the probability density of infectious agents

$$\partial_{tt}\iota(x,t) = \frac{2}{\tau}\left\{-\partial_t\iota(x,t) + \kappa_I k\sigma(x,t)\left[(1-p)(\iota(x,t)) + pI(t)\right] - \kappa_R\iota(x,t)\right\} + \kappa_I k(1-p)D_k\sigma(x,t)\partial_{xx}\iota(x,t), \tag{1}$$

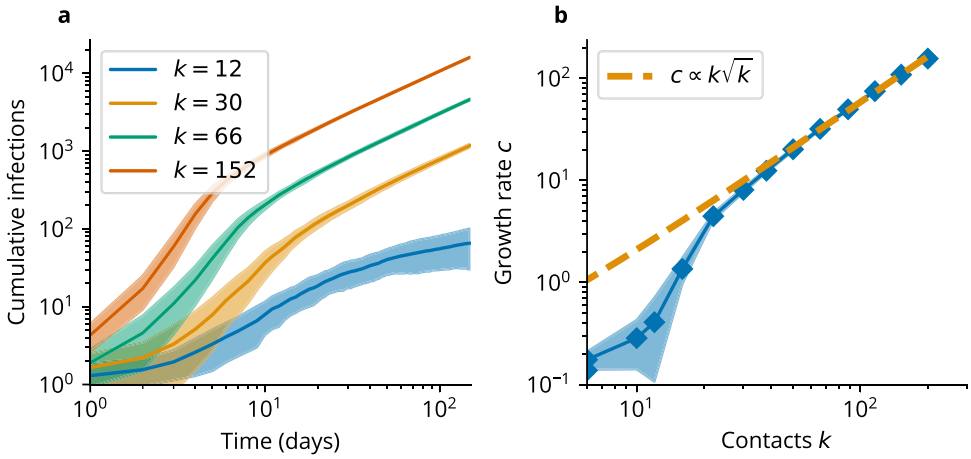

Figure 5. Comparison of model dynamics and mean-field approximation in the linear regime. (**a**) Cumulative number of infections in the model for a highly clustered network ($p = 0$). (**b**) Growth rate of cumulative infections in dependence of the number of contacts k. The growth rate scales as $c \propto k\sqrt{k}$ as predicted by mean-field theory for a large number of contacts k, see Eq. (2). Shown is the mean and standard error of the mean (SEM) of 20 independent simulations for each parameter.

where κ_I is the infection rate, and κ_R is the removal rate. Here, τ is the short time scale of the local disease dynamics of a single agent, while $D_k := \frac{\Delta x^2 \tilde{k}}{\tau}$, with $\tilde{k} := (k/2+1)(k+1)/12$ is a constant that determines the disease spread in the network on longer time scales. The total number of infectious agents I is defined as $I := \int \iota \, dx$, where the integral represents a nonlocal coupling by random contacts. The equation resembles the Telegrapher's equation but with a nonlocal coupling by random contacts and a nonlinear diffusion term due to local contacts. We recover the classical SIR model for $p = 1$, as expected. For the regime $p = 0$, we obtained the speed of disease spread through the network as

$$c = \sqrt{\kappa_I k D_k} \propto k\sqrt{k}. \tag{2}$$

Notably, the wave speed depends on k in a nonlinear manner. Comparing our prediction against simulation data (Supplementary Information) revealed that the wave speed is proportional to the growth rate of the cumulative infections in the linear regime, see Fig. 5.

Conclusion

We used Bayesian parameter inference to quantify the effects of government interventions in Germany on the transmission network of COVID-19 assuming it can be approximated by a Watts–Strogatz network. This network captures the key feature of social networks affected by NPIs, namely the strong clustering of contacts when people restrict their social life to a small, interconnected clique. Our analysis revealed that NPIs lead to a reduction in transmission probability, number of contacts, and, crucially, to the removal of almost all contacts outside highly clustered cliques. In contrast to standard epidemiological models, in this regime the cumulative number of infections does not increase exponentially but linearly, with a massively reduced peak of infections. The dynamics corresponds to a wave-like spread of the disease in the network, whose wave speed c we predicted by mean-field theory to scale as $c \propto k\sqrt{k}$ in dependence of the number of contacts k. At the same time, the effective reproduction number fluctuates around 1, irrespective of the wave speed. However, as long as the epidemic threshold is not reached by the reduction of contacts between cliques and the reduction of the infection probability, the disease still spreads in the population, which emphasizes the need for an effective test, trace, and isolate (TTI) system and a vaccinations.

Several other studies have highlighted the importance of random contacts between members of different cliques. A study that used mobility network data to show that the spread of the disease is mostly driven by infections at events which connect different communities, for example in restaurants and religious establishments, further supports the importance of random contacts[36]. Moreover, empirical studies of the circumstances under which people got infected revealed that although 46% to 66% of transmission is household-based (clustered contacts), random contacts between these cliques are essential to sustain the epidemic, even if they only cause a low percentage of infections directly[37]. A theoretical study that investigated the effect of different social network-based distancing strategies showed that the most effective social distancing strategy is to restrict contacts to a single clique, and to eliminate any contacts between the cliques (random contacts)[38].

We based our parameter inference on the daily count of positive tests. This approach is not perfect: the number of positive tests depends on the testing policy and the number of available tests. Also, the date, when the test result is recorded is always delayed. Another option is the usage of daily deaths, which are independent of testing. However, it is very difficult to base our approach on the death count, because the total death count in Germany during the first wave and especially during the summer was comparatively low. That means that even

our scaled-down system would be intractably large, if we want to reproduce the number of deaths using realistic assumptions about the fatality rate of the disease. Additionally, the time distribution between infection and death has a large variance of up to 10 days, as it varies depending on the age of the affected individual[29,39], increasing the noise in the already low death numbers. We further assumed that all people that were tested positive were also infectious at the time of the test and that the time delay between people turning infectious and their test is negligible. In future studies, the time delay between testing and the onset of infectiousness could be accounted for using nowcasting[40].

In our model we assumed a Watts–Strogatz transmission network and distinguished between random and clustered links. It is possible to use other algorithms to construct clustered networks, for example the stochastic block model[41], the relaxed caveman graph[42], or the configuration model with defined clustering[43], to name a few. We chose the Watts–Strogatz graph for its simplicity and its low number of parameters, but we expect that other graph ensembles would yield similar results, as long as they allow for high clustering and overlap of cliques. A potential target of future research could be the identification of the most appropriate network ensemble for different lockdown scenarios.

We did not account for the scale-free degree distribution found in real social networks. It is a result of hubs with a large number of connections potentially connecting several communities, like grocery stores, restaurants, religious establishments, etc. This favors disease spread even more than randomly assigned links and can lead to outbursts of infections (superspreading events)[26]. In fact, there were several superspreading events related to carnival festivities at the beginning of the first wave in Germany[32]. We believe that this, at least in part, explains why our parameter inference scheme finds larger values for the total number of social contacts k than previously reported[34] for the time period before NPIs were put in place. However, the fact that our inferred parameters for the later time periods are consistent with other reports regarding the number of contacts and the infection probability, reassures us that our assumption, that super spreading was less important after NPIs were enacted, is justified.

We inferred that lifting contact restrictions in Germany after June 6 moved the disease dynamics back into the exponential regime. However, the effective reproduction number remained close to one, due to a reduced transmission probability of the disease, compared to the pre-lockdown time period. This is likely a result of several phenomena: first, there is probably a small seasonal reduction in virus transmission due to the higher temperatures when no lockdown policy is in effect[44]. Second, even after contact restrictions were lifted, there was now a mask mandate at public places, which has been shown to reduce the daily growth rate of new cases by around 47% in Germany[45]. Lastly, mobility was still reduced by about 20% compared to previous years[30] correlating with lower virus transmission and indicating a higher awareness of the virus in the population[46,47]. In October, people did spend more time indoors again, and mobility reduced to normal levels, contributing to a second fast exponential growth of cases. The German government responded with new NPIs in November, which reduced the effective reproduction number to a value around one, again corresponding to a linear regime of new cases.

Our analysis shows that NPIs can reduce the effective reproduction number to one by eliminating random contacts. However, eliminating the disease does most likely require to cut almost all contacts between different cliques, for example, by working from home. In summary, government interventions should target random contacts and encourage people to restrict their contacts to a single clique in order to efficiently prevent disease spread.

Methods

Model definition. We study an agent-based, discrete-time SEIR model on the classical Watts–Strogatz network, representing agents as network nodes. The network is constructed by, first, connecting every node to its k nearest neighbors in a ring-like topology, and, second, rewiring every link to a random node with probability p, see Fig. 2. Every node i has a discrete state $s_i \in \mathcal{S} = \{S, E, I, R\}$, corresponding to susceptible (S), exposed (E), infectious (I), and removed (R) states.

Disease progression is dictated by Γ-distributed waiting times inferred from COVID-19 disease characteristics, as these have been found to describe the disease progression best[29]. During every discrete time step t, where the length of the time step corresponds to one day, each susceptible agent can become exposed with probability p_I for every infectious agent they are connected to,

$$P(s_i(t+1) = E | s_i(t) = S) = 1 - (1 - p_I)^{I_i}, \tag{3}$$

where I_i is the number of infectious agents node i is connected to. Upon infection, we change the agent's state to exposed, and assign a waiting time $\tau_E \sim \Gamma(k_E, \theta_E)$ which we draw from a Γ-distribution with shape k_E and scale θ_E. In every time step, the waiting times are reduced by one day, $\tau_E(t+1) = \tau_E(t) - 1$ if $\tau_E(t) > 0$. Else, the disease progresses, $s_i(t+1) = I$, and a new waiting time is assigned from another Γ-distribution $\tau_I \sim \Gamma(k_I, \theta_I)$, with shape k_I and scale θ_I. Finally, when $\tau_I \leq 0$, the node is removed, $s_i(t+1) = R$. A sketch of the SEIR dynamics can be found in Supplementary Fig. 1. See Table 1 for an overview of all model parameters.

We calculate the shape and scale of the Γ-distributions from the reported mean and variance of the time in the respective states according to

$$k_{E,I} = \frac{\langle \tau_{E,I} \rangle^2}{\langle \Delta\tau_{E,I}^2 \rangle}, \tag{4}$$

$$\theta_{E,I} = \frac{\langle \Delta\tau_{E,I}^2 \rangle}{\langle \tau_{E,I} \rangle}. \tag{5}$$

Symbol	Description
k	Mean number of contacts
p	Fraction of random contacts
p_I	Infection probability
$\langle \tau_{E,I} \rangle$	Mean exposed/infectious duration
$\langle \Delta \tau^2_{E,I} \rangle$	Variance of exposed/infectious duration

Table 1. Model parameters.

Note that we did not infer these parameters with our Bayesian parameter inference framework.

In our model we do not account for an inflow of infectious people by travel; we instead account for the initial surge of infections by placing randomly $n_E(0)$ exposed and $n_I(0)$ infectious agents in the population.

Bayesian parameter inference. We apply approximate Bayesian computation with a sequential Monte–Carlo scheme (ABC-SMC) to infer the set of parameters $\Theta = \{p_I, p, k, n_E(0), n_I(0)\}$ of our agent-based model. We always keep the total number of agents fixed at $n = 3 \cdot 10^5$. To this end, we employ the Python package pyABC[48]. In short, the algorithm employs sequential importance sampling over generations $T = 1, ..., n_T$. In generation T, the algorithm draws sets of parameters θ_i from a given proposal distribution and consequently simulates data $C^{(i)}$ from the model, until n_{ABC} instances were accepted based on the comparison to observed data via a distance function $D(C^{(i)}, C_{\text{obs}})$ and acceptance threshold ε_T, $D(C^{(i)}, C_{\text{obs}}) \leq \varepsilon_T$. As the distance function, we choose the absolute difference between new cases in the model instance $C^{(i)}(t)$ and the respective reports for Germany $C_{\text{obs}}(t)$,

$$D(C^{(i)}, C_{\text{obs}}) := \sum_t \left| C^{(i)}(t) - C_{\text{obs}}(t) \right|. \tag{6}$$

New cases in the model are given by the daily new infections

$$C^{(i)}(t) = n_I^{(i)}(t) - n_I^{(i)}(t-1) + n_R^{(i)}(t) - n_R^{(i)}(t-1). \tag{7}$$

They are compared to the 7-day rolling average of new case reports in Germany $C_{\text{obs}}(t)$ provided by Johns Hopkins University[28] to account for the weekly fluctuations in reporting. Acceptance of model instances depends on the acceptance threshold ϵ_T of generation T, which we choose as the median of the distances of the accepted instances of the previous generation

$$\epsilon_{T+1} = \text{median}\left(\left\{ D(C^{(i)}, C_{\text{obs}}) < \epsilon_T \right\} \right). \tag{8}$$

When n_{ABC} model instances have been accepted, the algorithm constructs new proposal distributions from the accepted instances to allow high acceptance rates while decreasing the threshold[49]. In particular, for the continuous variables it employs a multivariate normal distribution with an adaptive covariance matrix based on the sample covariance matrix, whose scale parameter is determined by a grid search with fivefold cross validation and refitting on the whole data set. We compute the discrete numbers of initially exposed and infectious people $n_E(0)$, $n_I(0)$ by rounding the continuous output of the multivariate normal distribution. We can do this without a large error as these parameters vary smoothly over a large range. For the parameter k we employ an adaptive discrete transition that assigns probabilities to all possible parameter values directly from the frequency of the respective value in the population of accepted particles with additional random jumps (with probability 0.3) to ensure absolute continuity of the prior. The process is repeated until the acceptance threshold is sufficiently low; we especially ensured that the threshold is considerably lower than the difference between the reported cases and the seven-day rolling average of cases.

Data availability

The simulation code and input data are available as Zenodo snapshot at https://doi.org/10.5281/zenodo.4884171.

Received: 31 May 2021; Accepted: 12 October 2021
Published online: 09 November 2021

References

1. Ferguson, N. M. *et al.* Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *Tech. Rep.* Imperial College London. https://doi.org/10.25561/77482 (2020).
2. van der Heiden, M. & Buchholz, U. Modellierung von Beispielszenarien der SARS-CoV-2-Epidemie 2020 in Deutschland [Modeling of example scenarios of the SARS-CoV-2 epidemic 2020 in Germany]. *Tech. Rep.* Robert Koch-Institut. https://doi.org/10.25646/6571.2 (2020).
3. Dehning, J. *et al.* Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* **369**, 9789. https://doi.org/10.1126/science.abb9789 (2020).
4. Khailaie, S. *et al.* Development of the reproduction number from coronavirus SARS-CoV-2 case data in Germany and implications for political measures. *BMC Med.* **19**, 1–16. https://doi.org/10.1186/s12916-020-01884-4 (2021).

5. Kühn, M. J. *et al.* Assessment of effective mitigation and prediction of the spread of SARS-CoV-2 in Germany using demographic information and spatial resolution. *Math. Biosci.* **339**, 108648. https://doi.org/10.1016/j.mbs.2021.108648 (2021).
6. Xiang, Y. *et al.* COVID-19 epidemic prediction and the impact of public health interventions: A review of COVID-19 epidemic models. *Infect. Dis. Model.* **6**, 324–342. https://doi.org/10.1016/j.idm.2021.01.001 (2021).
7. Britton, T., Ball, F. & Trapman, P. A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. *Science* **369**, 846–849. https://doi.org/10.1126/science.abc6810 (2020).
8. Kermack, W. O. & McKendrick, A. G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. Ser. A* **115**, 700–721. https://doi.org/10.1098/rspa.1927.0118 (1927).
9. Ridenhour, B., Kowalik, J. M. & Shay, D. K. Unraveling R_0: Considerations for public health applications. *Am. J. Public Health* **104**, e32–e41. https://doi.org/10.2105/AJPH.2013.301704 (2014).
10. Aleta, A. *et al.* Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav.* **4**, 964–971. https://doi.org/10.1038/s41562-020-0931-9 (2020).
11. Rader, B. *et al.* Crowding and the shape of COVID-19 epidemics. *Nat. Med.* **26**, 1829–1834. https://doi.org/10.1038/s41591-020-1104-0 (2020).
12. Lau, M. S. *et al.* Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. *Proc. Natl. Acad. Sci.* **117**, 22430–22435. https://doi.org/10.1073/pnas.2011802117 (2020).
13. Komarova, N. L., Schang, L. M. & Wodarz, D. Patterns of the COVID-19 pandemic spread around the world: exponential versus power laws. *J. R. Soc. Interface* **17**, 20200518. https://doi.org/10.1098/rsif.2020.0518 (2020).
14. Maier, B. F. & Brockmann, D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science* **368**, 742–746. https://doi.org/10.1126/science.abb4557 (2020).
15. Keeling, M. J. *et al.* Networks and the epidemiology of infectious disease. *Interdiscip. Perspect. Infect. Dis.* https://doi.org/10.1155/2011/284909 (2011).
16. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979. https://doi.org/10.1103/RevModPhys.87.925 (2015).
17. Watts, D. J. & Strogatz, S. H. Collective dynamics of small-world networks. *Nature* **393**, 440–442. https://doi.org/10.1038/30918 (1998).
18. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512. https://doi.org/10.1126/science.286.5439.509 (1999).
19. Herrmann, H. A. & Schwartz, J.-M. Why COVID-19 models should incorporate the network of social interactions. *Phys. Biol.* **17**, 065008. https://doi.org/10.1088/1478-3975/aba8ec (2020).
20. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203. https://doi.org/10.1103/PhysRevLett.86.3200 (2001).
21. Pastor-Satorras, R. & Vespignani, A. Epidemic dynamics in finite size scale-free networks. *Phys. Rev. E* **65**, 1–4. https://doi.org/10.1103/PhysRevE.65.035108 (2002).
22. Moreno, Y., Pastor-Satorras, R. & Vespignani, A. Epidemic outbreaks in complex heterogeneous networks. *Eur. Phys. J. B* **26**, 521–529. https://doi.org/10.1140/epjb/e20020122 (2002).
23. Thurner, S., Klimek, P. & Hanel, R. A network-based explanation of why most COVID-19 infection curves are linear. *Proc. Natl. Acad. Sci.* **117**, 22684–22689. https://doi.org/10.1073/pnas.2010398117 (2020).
24. Ball, F., Mollison, D. & Scalia-Tomba, G. Epidemics with two levels of mixing. *Ann. Appl. Probab.* **7**, 46–89 (1997).
25. Medo, M. Contact network models matching the dynamics of the COVID-19 spreading. *J. Phys. A Math. Theor.* **54**, 035601. https://doi.org/10.1088/1751-8121/abd34d (2021).
26. Small, M. & Tse, C. K. Small world and scale free model of transmission of SARS. *Int. J. Bifurcat. Chaos* **15**, 1745–1755. https://doi.org/10.1142/S0218127405012776 (2005).
27. Minter, A. & Retkute, R. Approximate Bayesian computation for infectious disease modelling. *Epidemics*. https://doi.org/10.1016/j.epidem.2019.100368 (2019).
28. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1 (2020).
29. Linton, N. *et al.* Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A statistical analysis of publicly available case data. *J. Clin. Med.* **9**, 538. https://doi.org/10.3390/jcm9020538 (2020).
30. Yilmazkuday, H. Stay-at-home works to fight against COVID-19: International evidence from Google mobility data. *J. Hum. Behav. Soc. Environ.* **31**, 210–220. https://doi.org/10.1080/10911359.2020.1845903 (2021).
31. Jing, Q.-L. *et al.* Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: A retrospective cohort study. *Lancet Infect. Dis.* **20**, 1141–1150. https://doi.org/10.1016/S1473-3099(20)30471-0 (2020).
32. Streeck, H. *et al.* Infection fatality rate of SARS-CoV-2 infection in a German community with a super-spreading event. *Nat. Commun.* **11**, 1–12. https://doi.org/10.1101/2020.05.04.20090076 (2020).
33. Chen, X. *et al.* Serological evidence of human infection with SARS-CoV-2: A systematic review and meta-analysis. *Lancet Glob. Health* **9**, e598–e609. https://doi.org/10.1016/S2214-109X(21)00026-7 (2021).
34. Mossong, J. *et al.* Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74. https://doi.org/10.1371/journal.pmed.0050074 (2008).
35. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342. https://doi.org/10.1126/science.1245200 (2013).
36. Chang, S. *et al.* Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* **589**, 82–87. https://doi.org/10.1038/s41586-020-2923-3 (2021).
37. Lee, E. C., Wada, N. I., Grabowski, M. K., Gurley, E. S. & Lessler, J. The engines of SARS-CoV-2 spread. *Science* **370**, 406–407. https://doi.org/10.1126/science.abd8755 (2020).
38. Block, P. *et al.* Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nat. Hum. Behav.* **4**, 588–596. https://doi.org/10.1038/s41562-020-0898-6 (2020).
39. Faes, C. *et al.* Time between symptom onset, hospitalisation and recovery or death: Statistical analysis of Belgian COVID-19 patients. *Int. J. Environ. Res. Public Health* **17**, 7560. https://doi.org/10.3390/ijerph17207560 (2020).
40. Höhle, M. & van der Heiden, M. Bayesian nowcasting during the STEC O104:H4 outbreak in Germany, 2011. *Biometrics* **70**, 993–1002. https://doi.org/10.1111/biom.12194 (2014).
41. Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic blockmodels: First steps. *Soc. Netw.* **5**, 109–137. https://doi.org/10.1016/0378-8733(83)90021-7 (1983).
42. Fortunato, S. Community detection in graphs. *Phys. Rep.* **486**, 75–174. https://doi.org/10.1016/j.physrep.2009.11.002 (2010).
43. Newman, M. E. J. Random graphs with clustering. *Phys. Rev. Lett.* **103**, 058701. https://doi.org/10.1103/PhysRevLett.103.058701 (2009).
44. Smith, T. P. *et al.* Temperature and population density influence SARS-CoV-2 transmission in the absence of nonpharmaceutical interventions. *Proc. Natl. Acad. Sci.* **118**, e2019284118. https://doi.org/10.1073/pnas.2019284118 (2021).
45. Mitze, T., Kosfeld, R., Rode, J. & Wälde, K. Face masks considerably reduce COVID-19 cases in Germany. *Proc. Natl. Acad. Sci.* **117**, 32293–32301. https://doi.org/10.1073/pnas.2015954117 (2020).

46. James, N. & Menzies, M. Efficiency of communities and financial markets during the 2020 pandemic. *Chaos* **31**, 083116. https://doi.org/10.1063/5.0054493 (2021).
47. Zachreson, C. *et al.* Risk mapping for COVID-19 outbreaks in Australia using mobility data. *J. R. Soc. Interface* **18**, 20200657. https://doi.org/10.1098/rsif.2020.0657 (2021).
48. Klinger, E., Rickert, D. & Hasenauer, J. pyABC: Distributed, likelihood-free inference. *Bioinformatics* **34**, 3591–3593. https://doi.org/10.1093/bioinformatics/bty361 (2018).
49. Jagiella, N., Rickert, D., Theis, F. J. & Hasenauer, J. Parallelization and high-performance computing enables automated statistical inference of multi-scale models. *Cell Syst.* **4**, 194–206. https://doi.org/10.1016/j.cels.2016.12.002 (2017).

Acknowledgements

We thank Prof. Jan Hasenauer for valuable discussions and input about Bayesian parameter inference and Prof. Michael Meyer-Hermann for helpful feedback on the manuscript. We thank the Centre for Information Services and High Performance Computing at Technische Universität Dresden for providing high-performance computing infrastructure. Diana David-Rus thanks Prof. Manfred Wildner and the Department of Infectious Diseases Epidemiology Surveillance and the Task Force at LGL-Bavaria for their support in ensuring the appropriate infrastructure. S.S. is supported by the European Social Fund (ESF) (100327771), co-financed by tax funds on the basis of the budget adopted by the members of the Saxon State Parliament. Y.S. is supported by the German Federal Ministry of Education and Research (FitMultiCell; Grant Number 031L0159A). H.H. is supported by the Volkswagen Foundation within the *Life?* program (Grant Number 96732). H.H. also has received funding from the Bundesministeriums für Bildung, und Forschung (BMBF) under grant agreement No. 031L0237C (MiEDGE project/ERACOSYSMED).

Author contributions

S.S.: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Writing—Original Draft, Visualization. D.D.-R.: Resources. Y.S.: Software. H.H.: Conceptualization, Methodology, Resources. A.D.: Conceptualization, Methodology, Supervision. All authors reviewed the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-01407-y.

Correspondence and requests for materials should be addressed to A.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021