Heterogeneity is essential for contact tracing

Bjarke Frost Nielsen,[1] Kim Sneppen,[1] Lone Simonsen,[2] and Joachim Mathiesen[1]

[1] *Niels Bohr Institute, University of Copenhagen, 2100 Copenhagen, Denmark*

[2] *Department of Science and Environment, Roskilde University, 4000 Roskilde, Denmark*

(Dated: June 5, 2020)

Background: The COVID-19 pandemic is most often modelled by well-mixed models, sometimes stratified by age and work. People are, however, different from one another in terms of interaction frequency as well as in formation of social groups. This contact heterogeneity especially challenges models of contact tracing (CT), but also predictions of epidemic severity generally. We explore how heterogeneity affects CT effectiveness and overall epidemic severity, using a real-world contact network.

Methods: Utilizing smartphone proximity data from Danish university students, we simulate the spread of COVID-19 on a network with realistic contact structure. Two modes of network homogenization are implemented to probe effects of heterogeneity. We then simulate a CT scheme on the network and explore the impact of heterogeneity, testing probability and contact threshold for quarantining.

Results: Measuring contact heterogeneity, we find an exponential distribution which persists on a timescale of several weeks. Comparing the true network to edge-swapped and randomized versions, we find that heterogeneity decreases the severity of COVID-19 in general, and that it drastically improves CT.

Conclusions: To capture heterogeneity, it is necessary to reconsider disease transmission models. Our findings show that heterogeneity is essential for CT, and that CT is effective even if only the most frequent contacts can be tracked down. We find that contact heterogeneity impedes the spread of COVID-19 in comparison with well-mixed networks. In perspective, this means that fitting traditional SEIR models to epidemic data is likely to overestimate the severity.

INTRODUCTION

Realistic contact and proximity data is hard to come by, and the health authorities of many nations have relied on well-mixed compartmental models of disease spreading to model the current COVID-19 epidemic [1, 2]. While the appeal of these models lies in their simplicity, they are generally not adequate in cases where features of *contact networks* are important. Furthermore, it is common to partition the populations of such models into sub-populations according to e.g. age groups and occupational, home and social spheres [3–8]. However, even if one may assign reduced interaction rates between certain groups, the well-mixed nature of the models fails to capture important features of restricted interactions [8].

Realistic interactions are characterized by a degree of monotony; you meet the same people regularly. In a well-mixed model – even if stratified by e.g. age – your contacts are essentially drawn at random at each new instant. Furthermore, individuals vary in their overall number of contacts and form groups based on social preferences – phenomena which contribute to transmission heterogeneity. In some previous epidemics, perhaps most notably the Ebola epidemic of 2014-16, transmission heterogeneity and spatial dynamics have been shown to be of tremendous importance [9].

Lately, contact tracing – a mitigation strategy which relies directly on contact network structure – has been the center of much attention due to its promises of epidemic control in a relatively open society [10–15]. With interventions such as this, realistic contact networks are indispensable, and the usual well-mixed approach is insufficient – even more so than when modelling unmitigated spreading [16, 17].

In this paper, we utilize Bluetooth proximity data obtained from a cohort of university students at a large European university (see *Methods* for details). These data are similar in nature to the sort of readings one might obtain from contact tracing smartphone applications [18], meaning that they provide a useful virtual laboratory for contact tracing. While the data only comprise a section of the total contact network of each participant, they display well-defined heterogeneity, the effects of which can be studied, and compared with analogous homogenized networks. The participant group is homogeneous in age and occupation, and would consequently usually be modelled as homogeneous – a modelling assumption that we can directly probe the validity of, in the context of contact tracing.

The first part of the paper deals with the effects of **contact heterogeneity** on the spread of COVID-19. Three degrees of heterogeneity are introduced: **i)** the unaltered, realistic network. **ii)** an edge-swapped version of the network [19], retaining contact heterogeneity but eliminating group formation preferences, including spatial preferences. **iii)** a randomized network, retaining only the overall (mean) contact frequency, but eliminating heterogeneity. This allows us to investigate questions such as whether it affects the spread of COVID-19 if some people are more social than others, and whether network structure such as group formation are important for the spread of COVID-19.

In the second part, we introduce a **contact tracing**

scheme and simulate this on the network. This allows us to address the influence of contact heterogeneity *on* contact tracing – is a fairly well mixed version of the contact network just as *traceable*? Do spatial effects and group formation affect contact tracing? Furthermore, we explore two key parameters which influence the effectiveness of contact tracing, namely the probability of testing and the contact threshold – the threshold duration of contact with an exposed individual required to quarantine. The latter is especially interesting, since it is a directly controllable parameter when e.g. designing contact tracing smartphone applications [18].

METHODS

We analyze social proximity and contact dynamics from data collected by smartphones distributed among around 1000 participants (undergraduate students at the Technical University of Denmark [20, 21]). The smartphones were equipped with an application that collected communication in the form of call and text messaging logs, geolocation of the users by records of GPS coordinates and social proximity using the Bluetooth port. All smartphones in the study were programmed to synchronously open their Bluetooth ports every 5 minutes to scan for nearby devices included in the study and to record the GPS coordinates. The data we consider have been collected over a period of two years, 2013-2015.

The distance between participants is inferred from the strength (RSSI) of the Bluetooth signal being sent between the devices. The signal strength can resolve distances in the range of ≤ 1 meter to approximately 10-15 meters [22]. We define a *contact* between two individuals whenever the Bluetooth signal strength between their respective devices exceeds -85dBm. This definition of contact captures essentially all ≤ 1m interactions while excluding a large portion of the 3m interactions and above [22], in line with the recommendations of public health authorities [23, 24]. With this definition of contact we can create a well-defined time-dependent contact network where individuals are represented by nodes and social contact by time-dependent links, similarly to the work of [25]. The link activity, or the contact between nodes, is resolved in temporal windows of 5 minutes. The time-dependent contact network will be the basis for our modelling of the transmission of COVID-19.

Modelling COVID-19 on the network We model the spread of COVID-19 in the population of the study by an agent-based model (where the study participants serve as the agents) with five states: **S**usceptible to the disease, **E**xposed, **P**re-symptomatic (but infectious), **I**nfected (potentially with symptoms) and **R**ecovered/Removed. In the absence of contact tracing (described below), the P and I states are identical, in that an individual in one of these states can infect others. Aside from these mutually exclusive states, persons can also be flagged as **Q**uarantined. This only comes into play during our contact tracing simulations, as described below.
When a susceptible person comes into contact with a person in the **I** or **P** state, there is a probability p_{inf} of transmission in each 5-minute window. The basic model (sans contact tracing) thus has four parameters: Transmission probability upon contact p_{inf}, a time-scale characterizing the exposed state τ_{E}, a time-scale characterizing the presymptomatic state τ_{P} and a time-scale characterizing the infected state τ_{I}.

As shown in the model illustration of Fig. 2, the incubation time is Γ-distributed with a mean of 3.6 days, of which 1.2 days comprise the presymptomatic infectious state. The infectious state, where symptoms may be displayed, is then 5 days. The last parameter of the basic model, the transmission probability in each window of time, is fitted to reproduce a daily growth rate of 23% in the early epidemic, based on estimates from [26].

In our simulations we initiate the epidemic by randomly assigning 2% of the population to be infected, and follow the evolution of the epidemic on the real contacts in Fig. 1 and Fig. 3 (the solid curves). By employing two different modes of *shuffling* of network connections (edges), we further study the effects of contact heterogeneity on the one hand, and space and group formation on the other. The first method, *edge swapping*, preserves the degree of connectivity of each person (node), while destroying any spatial and group formation preference [19]. The second method, *randomization*, preserves only the overall connectivity level in each window of time, but homogenizes the number of contacts for each person. In Fig. 3 we plot the epidemic trajectories by averaging over 20 simulations.

Simulating contact tracing The entire scheme around contact tracing consists of two parts: *regular testing* of symptomatic individuals (with a testing probability $p_{\text{test}} < 1$) and the contact tracing algorithm itself, which is activated once an individual tests positive. Once a positive individual is found by *regular testing*, their recent contacts are put in quarantine for a specified time (5 days as suggested by [15]) and tested once the quarantine period has elapsed (before potential release).
In other words, the contact tracing scheme proceeds as follows:

- Contacts of the positive individual are traced, with only relatively recent contacts being retained. In the simulations presented here, only contacts encountered up to 5 days before symptom onset are kept.

- If the total duration of close proximity exceeds the *contact threshold*, the contact is quarantined for 5 days.

- After the quarantine period has elapsed, the individual is tested. If negative, the individual is released. Otherwise a new 5-day quarantine is issued.

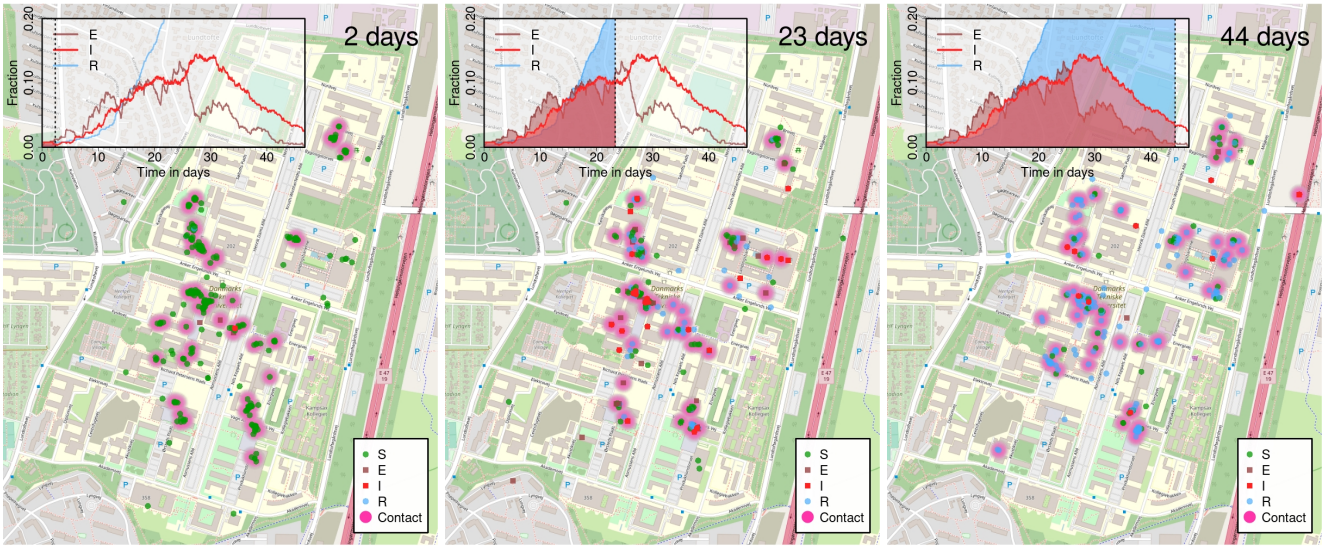

FIG. 1. Simulating the spread of COVID-19 on the contact network. Here, the approximate geographical positions of individuals are shown on the DTU Campus for a representative run of the model. Regions of contact (defined by signal strength exceeding the −85dBm cutoff) are shown as diffuse clouds of pink. Snapshots shown are at day 2, 23 and 44 of the outbreak.

- The procedure is iterated again for any contacts of the quarantined person, if this individual is tested positive.

The quarantine flag is handled simply, by assuming perfect efficiency, meaning that any quarantined person is temporarily unable to contact others.
The two simplest assumptions regarding the *regular testing* scheme are that symptomatic individuals are tested at a constant *rate* (throughout their illness), *or* that they have a *fixed probability* of being tested when first developing symptoms. We have compared the two and found that they perform comparably (see Supporting Information), and thus we work with the fixed probability scheme here, for simplicity.

RESULTS

Contact heterogeneity. The distribution of daily contact events in Fig. 2 shows a marked contact heterogeneity, characterized by an exponential shape (see Fig. 2b) with a coefficient of variation of 1.03 and a mean of 131. Even more importantly, a significant degree of contact heterogeneity is retained, albeit with some attenuation, when exploring an entire 7-week window. Here the coefficient of variation is 0.95, still close to the value for an exponential distribution, and the mean is 86. It is clear that extremely social behaviour becomes less frequent over this longer time-window, reflecting that individuals do not come into university every day, over this longer period. However, the significant degree of contact heterogeneity still present shows that it approximately represents a *quenched disorder*, which affects the entire epidemic trajectory and does not simply average out over the course of an epidemic. In the following we explore the often profound consequences of this finding.

Heterogeneity reduces epidemic severity. In Fig. 3, we show the simulated evolution of COVID-19 on three different contact networks: The true network (unshuffled), the edge-swapped and the fully randomized network where each person is assigned an average contact frequency. Each trajectory is averaged over 20 runs, each similar in nature to the one shown in the inserts of Fig. 1.

The overall findings are that:

- The total number of exposed individuals is very sensitive to contact heterogeneity, but *not* to spatial effects and social preferences.

  Contact heterogeneity evidently *prevents* the disease from spreading to all parts of the network, with the total fraction exposed reaching 71% in the true network and 94% in the randomized network.

- The infection peak, on the other hand, is sensitive to spatial effects *as well as* contact heterogeneity.

  As such, the peak load increases by some 5 percentage points when spatial structure is destroyed, and by 9 percentage points when contact structure is homogenized as well.

- Overall, the group formation and spatial structure only has the effect of slowing the progression somewhat, but doesn't affect the attack rate.

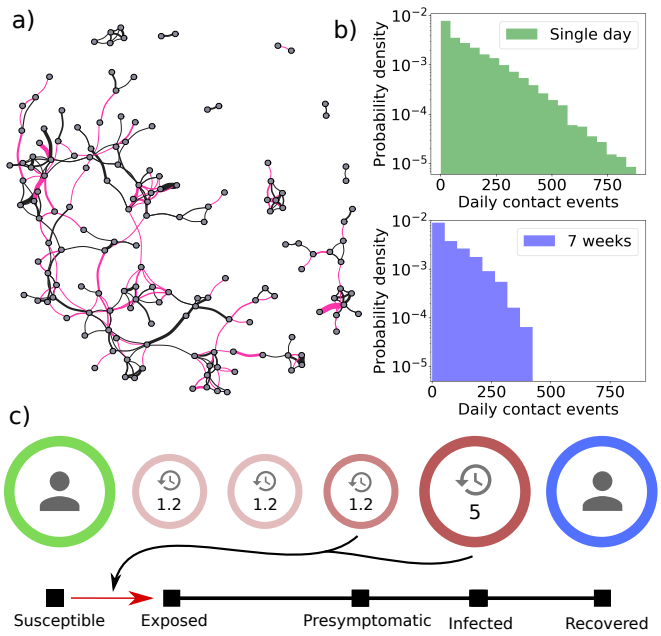

FIG. 2. **a)** A small subset of a contact network for one week. Link thickness indicates cumulated contact time, and a lower cut-off of approximately 2 hours is chosen. Black links are repeated, relative to the previous week, while red links are new. **b)** Top: Histogram of contact events over a single day (logarithmic plot). The coefficient of variation is $c_V = 1.03$ and the mean is $\mu = 131$. Bottom: Histogram of contact events over seven weeks, scaled to a daily rate (logarithmic plot). Here, $c_V = 0.95$ and $\mu = 86$. Both plots show a marked heterogeneity, demonstrating that contact heterogeneity is approximately a quenched disorder on the timescale of a few weeks. **c)** Our agent-based model of COVID-19 spreading on a contact network. Individuals in the **S**usceptible state may be exposed by those in the **P**resymptomatic as well as **I**nfected states. The Exposed-Presymptomatic triplet of states together comprise the Γ-distributed incubation period.

Repeated contacts are essential for contact tracing. The effects of contact heterogeneity and social preference on contact tracing can be probed by utilizing the same two modes of network re-shuffling as in the previous section. In Fig. 4 we compare how the contact tracing algorithm performs on the three networks. It is clear that the epidemic trajectory of the *true* network is drastically altered by contact tracing, with the attack rate and infection peak being profoundly attenuated. In the two shuffled networks, on the other hand, we see very little benefit from contact tracing. This leads to the conclusion, that contact tracing *depends* on social preference and contact heterogeneity.

Recently, several studies have found significant heterogeneity in COVID-19 transmission [27, 28]. Relatedly, it was shown in an agent-based model that heterogeneity in *infectiousness* has a considerable impact on the feasibility of COVID-19 mitigation strategies [29]. Our results show that another type of heterogeneity, namely the social kind, has a similarly profound effect on the

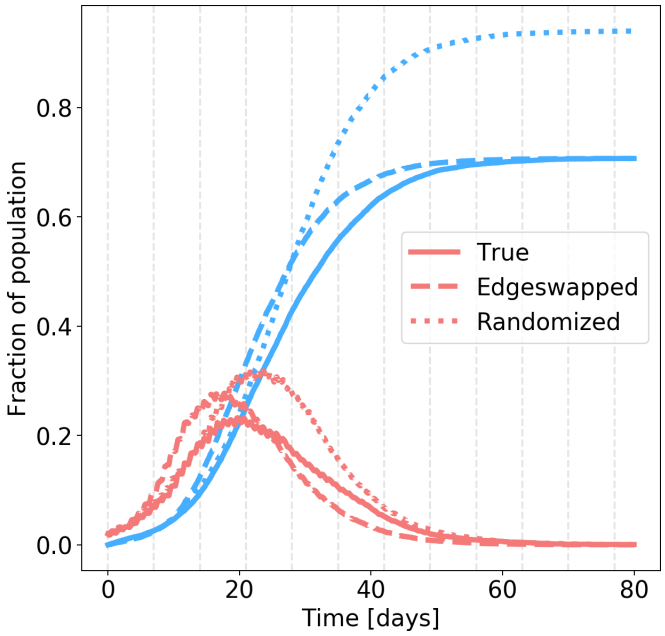

FIG. 3. Comparison of exposed + presymptomatic + infected (**red**) and recovered (**blue**) individuals in the three networks types. *True network* (full lines): Total fraction exposed: 71%. Infection peak: 28%. *Edgeswapped network* (dashed lines): Total fraction exposed: 71%. Infection peak: 33%. c) *Randomized network* (dotted lines): Total fraction exposed: 94%. Infection peak: 37%. The infection probability per 5-minute contact is $p_{\text{inf}} = 0.5\%$, fitted to produce a daily growth rate of 23% in the early epidemic, for the true network.

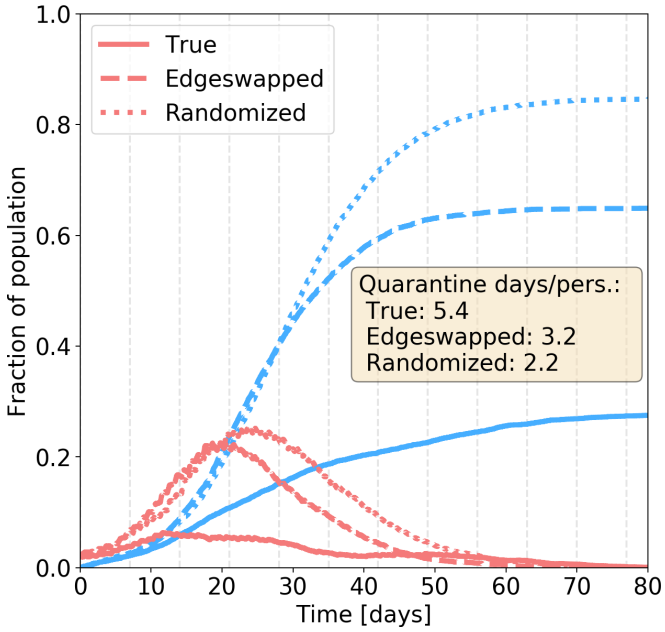

FIG. 4. **Contact tracing.** Comparison of exposed + presymptomatic + infected (**red**) and recovered (**blue**) individuals in the three networks types. Disease parameters are identical to those of Fig. 3. The contact threshold for quarantining is approximately 2 hours (125 minutes) while the testing probability is set at 25%.

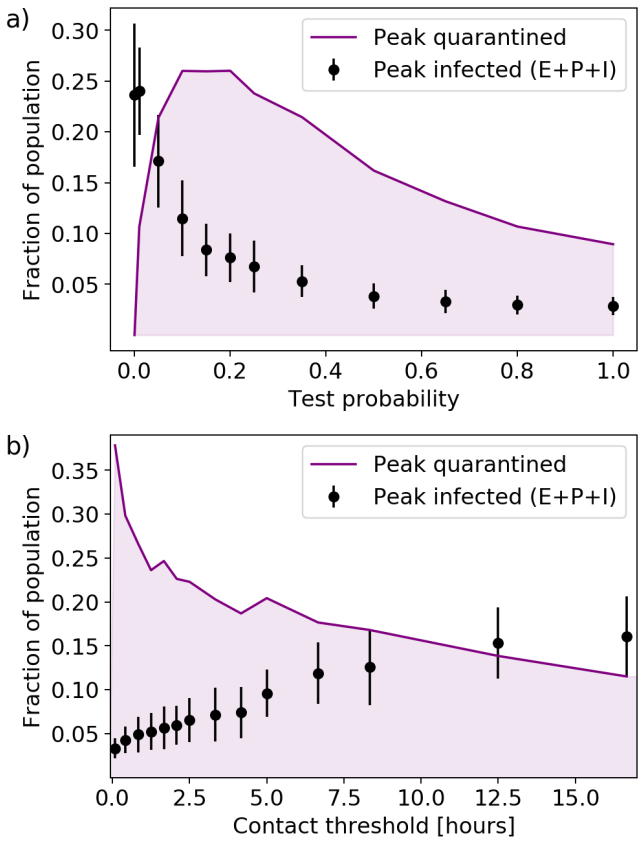

FIG. 5. Contact tracing effectiveness. Disease parameters are identical to those of Fig. 3. a) Testing probability vs peak infection and quarantine fraction. The contact threshold is set at 125 minutes. b) Contact threshold vs peak infection and quarantine fraction. The testing probability is set at 25%. For each value of the parameter, 50 simulations were run. The error bars indicate the standard deviation of outcomes of individual simulations.

effectiveness of contact tracing.

Estimating an optimal contact threshold for efficient tracing. When making public health decisions about COVID-19 mitigation schemes, it is first and foremost important to have reliable predictions regarding the overall magnitude of effects from each considered strategy. Next, however, it is advantageous to identify the parameters which influence the effectiveness. Some of these are bound to be out of our control – properties which are intrinsic to SARS-CoV-2 – while others can be influenced, or even constitute design decisions on our part.

In the two following sections, we explore two central parameters, namely the *testing probability* and *contact threshold*. The former determines the probability of being tested if sick, while the latter determines how readily quarantines are issued when contact with an infectious person has been established.

The testing probability: The regular testing required in contact tracing is determined by a *testing probability* which reflects several factors, such as general testing availability, symptom development and willingness to participate in testing. In Fig. 5a, we explore the influence of the testing probability on the peak load in terms of quarantined and exposed individuals.

Unsurprisingly, the quarantine fraction vanishes at very low testing probabilities, and the infection peak attains its maximal value. While the infection load is a decreasing function of testing, the quarantine fraction does not display a simple monotonic response to an increase in testing. Rather, it attains a maximum around 10%, followed by a gradual decline. This clearly shows that changes in testing availability should go hand-in-hand with considerations of the influence on the quarantine fraction, and that the relation is nontrivial.

The contact threshold: When performing contact tracing, it is necessary to define a *contact threshold*, meaning the minimum duration of proximity between an infectious and a susceptible person which results in quarantine. A low contact threshold thus intuitively means that a large fraction of contact persons will be placed in quarantine, when a positive individual is found. In Fig. 5b, the contact threshold is shown to have a profound effect on the infection peak as well as the peak fraction of the population in quarantine. As intuition would have it, the infection peak is clearly an increasing function of the contact threshold, while the quarantine fraction decreases.

Above a contact threshold of approximately two hours of cumulative proximity (contact in 25 5-minute windows), the quarantine fraction decreases only slowly. The peak infection load, on the other hand, increases steadily with the threshold, with a reduction of the epidemic peak from 25% without contact tracing to 8% when quarantining only contacts of at least 4 hours accumulated duration within a 5 day window. This goes to show that contact tracing is effective, *even if* it is only possible to locate frequent contacts (and 25% of the infected people).

DISCUSSION

This paper explores the effect of contact heterogeneity on the dynamics and mitigation of an epidemic. Remarkably we find that different people have quite different social activity, and further that this activity was well approximated by an exponential distribution (see Fig. 2). This observation is not far from the findings of [3], where a coefficient of variation of social contacts of about 0.8 was reported for people between 20 and 30 years. Further we found that this person-specific social activity remained constant over long time intervals, as seen by the fact that both the one day, and the 7 week activity pattern had coefficients of variations close to 1. Thus the social activity is approximately fixed to each person, and

represents a quenched disorder that influences the entire epidemic trajectory.

We find that the quenched disorder of contacts in the system, significantly impedes the spread of the disease and that it makes mitigation by contact tracing more efficient. Thus it is clear that traditional well-mixed S(E)IR models would overestimate the severity of the epidemic, or, conversely, lead to an underestimation of transmission risk, when fitting to measurements. In a previous modelling study [30], it was shown that heterogeneity in the *susceptibility* of individuals likewise reduces the overall severity.

Most importantly, we found that contact heterogeneity has a profound influence on the effectiveness of contact tracing (Fig. 4): A higher degree of contact heterogeneity make mitigation by tracing much more cost effective. Correspondingly, models which neglect heterogeneity and social clustering are likely to underestimate the feasibility of contact tracing schemes. Indeed, our modelling suggests that contact tracing has the potential to be an effective mitigation strategy for COVID-19.

We also explored the effect of two central parameters, the *testing probability* and the *contact threshold*, on a simple contact tracing scheme.

The testing probability is influenced by some factors which are within our control, such as the overall availability of testing, as well as some factors which are essentially intrinsic to SARS-CoV-2, such as the development of symptoms. Testing probability has a non-trivial relation to quarantine fraction, with a peak in quarantined individuals occurring at around 10% probability.

The contact threshold, on the other hand, is a controllable parameter and essentially constitutes a design decision when e.g. developing contact tracing applications, such as the framework by Google and Apple [18]. We find that contact tracing is effective, even if only relatively frequent contacts are quarantined: Even limiting testing and quarantine to only include secondary contacts with more that 4 hours exposure within a 5 day period leads to a substantial decrease in epidemic severity. Notably, this was done in a simulation where the whole epidemic develops in a university environment and thus does not include all social contacts. In line with the fact that a person infected with COVID-19 infects between 0.5 and 1 person per day, on average, we therefore speculate that real contact tracing apps would work with even larger threshold for test and quarantine.

[1] Norwegian Institute of Public Health, Coronavirus modelling at the NIPH, `https://www.fhi.no/en/id/infectious-diseases/coronavirus/coronavirus-modelling-at-the-niph-fhi/` (2020), [Online; accessed 28-May-2020].

[2] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, Epidemic analysis of covid-19 in china by dynamical modeling (2020), arXiv:2002.06563 [q-bio.PE].

[3] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, *et al.*, Social contacts and mixing patterns relevant to the spread of infectious diseases, PLoS medicine **5** (2008).

[4] P. Klepac, S. Kissler, and J. Gog, Contagion! the bbc four pandemic–the model behind the documentary, Epidemics **24**, 49 (2018).

[5] P. Klepac, A. J. Kucharski, A. J. Conlan, S. Kissler, M. Tang, H. Fry, and J. R. Gog, Contacts in context: large-scale setting-specific social mixing matrices from the bbc pandemic project, medRxiv (2020).

[6] L. Pellis, S. Cauchemez, N. M. Ferguson, and C. Fraser, Systematic selection between age and household structure for models aimed at emerging epidemic predictions, Nature communications **11**, 1 (2020).

[7] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. Pearson, J. D. Munday, *et al.*, The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study, The Lancet Public Health (2020).

[8] M. J. Keeling, E. Hill, E. Gorsich, B. Penman, G. Guyver-Fletcher, A. Holmes, T. Leng, H. McKimm, M. Tamborrino, L. Dyson, and M. Tildesley, Predictions of covid-19 dynamics in the uk: short-term forecasting and analysis of potential exit strategies, medRxiv 10.1101/2020.05.10.20083683 (2020).

[9] M. S. Lau, B. D. Dalziel, S. Funk, A. McClelland, A. Tiffany, S. Riley, C. J. E. Metcalf, and B. T. Grenfell, Spatial and temporal dynamics of superspreading events in the 2014–2015 west africa ebola epidemic, Proceedings of the National Academy of Sciences **114**, 2337 (2017).

[10] M. J. Keeling, T. D. Hollingsworth, and J. M. Read, The efficacy of contact tracing for the containment of the 2019 novel coronavirus (covid-19)., medRxiv 10.1101/2020.02.14.20023036 (2020).

[11] R. M. Anderson, H. Heesterbeek, D. Klinkenberg, and T. D. Hollingsworth, How will country-based mitigation measures influence the course of the covid-19 epidemic?, The Lancet **395**, 931 (2020).

[12] J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, W. J. Edmunds, F. Sun, *et al.*, Feasibility of controlling covid-19 outbreaks by isolation of cases and contacts, The Lancet Global Health (2020).

[13] A. J. Kucharski, P. Klepac, A. Conlan, S. M. Kissler, M. Tang, H. Fry, J. Gog, J. Edmunds, C. C.-. W. Group, *et al.*, Effectiveness of isolation, testing, contact tracing and physical distancing on reducing transmission of sars-cov-2 in different settings, medRxiv (2020).

[14] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dorner, M. Parker, D. G. Bonsall, and C. Fraser, Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing, medRxiv 10.1101/2020.03.08.20032946 (2020).

[15] A. Eilersen and K. Sneppen, Estimating cost-benefit of quarantine length for covid-19 mitigation, medRxiv

10.1101/2020.04.09.20059790 (2020).

[16] M. Gasparek, M. Racko, and M. Dubovsky, A stochastic, individual-based model for the evaluation of the impact of non-pharmacological interventions on covid-19 transmission in slovakia, medRxiv 10.1101/2020.05.11.20096362 (2020).

[17] A. Aleta, D. Martin-Corral, A. Pastore y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini, S. Merler, A. Pentland, A. Vespignani, E. Moro, and Y. Moreno, Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the covid-19 epidemic, medRxiv 10.1101/2020.05.06.20092841 (2020).

[18] Apple Inc., Building an App to Notify Users of COVID-19 Exposure, `https://developer.apple.com/documentation/exposurenotification/building_an_app_to_notify_users_of_covid-19_exposure` (2020), [Online; accessed 31-May-2020].

[19] S. Maslov and K. Sneppen, Specificity and stability in topology of protein networks, Science **296**, 910 (2002).

[20] A. Stopczynski, V. Sekara, P. Sapiezynski, A. Cuttone, M. M. Madsen, J. E. Larsen, and S. Lehmann, Measuring large-scale social networks with high resolution, PloS one **9**, e95978 (2014).

[21] A. Mollgaard, I. Zettler, J. Dammeyer, M. H. Jensen, S. Lehmann, and J. Mathiesen, Measure of node similarity in multilayer networks, PloS one **11** (2016).

[22] V. Sekara and S. Lehmann, The strength of friendship ties in proximity sensor data, PLOS ONE **9**, 1 (2014).

[23] World Health Organization, Q&A on coronaviruses (COVID-19), `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses` (2020), [Online; accessed 28-May-2020].

[24] Centers for Disease Control and Prevention, How COVID-19 Spreads, `https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html` (2020), [Online; accessed 28-May-2020].

[25] A. Stopczynski, A. S. Pentland, and S. Lehmann, Physical proximity and spreading in dynamic social networks, arXiv preprint arXiv:1509.06530 (2015).

[26] Our World In Data and European Centre for Disease Prevention and Control, covid-19-data (Deaths), `https://github.com/owid/covid-19-data/` (2020).

[27] Y. Zhang, Y. Li, L. Wang, M. Li, and X. Zhou, Evaluating transmission heterogeneity and super-spreading event of covid-19 in a metropolis of china, International Journal of Environmental Research and Public Health **17**, 3705 (2020).

[28] B. M. Althouse, E. A. Wenger, J. C. Miller, S. V. Scarpino, A. Allard, L. Hébert-Dufresne, and H. Hu, Stochasticity and heterogeneity in the transmission dynamics of sars-cov-2 (2020), arXiv:2005.13689 [q-bio.PE].

[29] K. Sneppen and L. Simonsen, Impact of superspreaders on dissemination and mitigation of covid-19, medRxiv 10.1101/2020.05.17.20104745 (2020).

[30] M. G. M. Gomes, R. M. Corder, J. G. King, K. E. Langwig, C. Souto-Maior, J. Carneiro, G. Goncalves, C. Penha-Goncalves, M. U. Ferreira, and R. Aguas, Individual variation in susceptibility or exposure to sars-cov-2 lowers the herd immunity threshold, medRxiv 10.1101/2020.04.27.20081893 (2020).