SARS-CoV-2 superspreading in cities vs the countryside

ANDREAS EILERSEN and KIM SNEPPEN

Niels Bohr Institute, University of Copenhagen, København Ø, Denmark

Eilersen A, Sneppen K. SARS-CoV-2 superspreading in cities vs the countryside. APMIS. 2021; 129: 401–407.

The first wave of the COVID-19 pandemic was characterized by an initial rapid rise in new cases followed by a peak and a more erratic behaviour that varies between regions. This is not easy to reproduce with traditional SIR models, which predict a more symmetric epidemic. Here, we argue that superspreaders and population heterogeneity would predict such behaviour even in the absence of restrictions on social life. We present an agent-based lattice model of a disease spreading in a heterogeneous population. We predict that an epidemic driven by superspreaders will spread rapidly in cities, but not in the countryside where the sparse population limits the maximal number of secondary infections. This suggests that mitigation strategies should include restrictions on venues where people meet a large number of strangers. Furthermore, mitigating the epidemic in cities and in the countryside may require different levels of restrictions.

Key words: COVID-19; model; superspreading; population density; epidemiology.

Kim Sneppen, Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 København Ø, Denmark. e-mail: sneppen@nbi.ku.dk

At its onset the coronavirus disease 2019 (COVID-19) pandemic shocked the world, with the number of new cases and deaths growing more than 20% per day in the main hotspots [1]. With a growth rate this high, the disease was expected to spread through the population in less than six months without mitigation, and to reach a peak after three months, at which point 30% of the population would have had the disease [2]. This, however, was not how the initial wave of the epidemic played out [3].

While most of the epidemic undoubtedly was halted due to mitigation efforts, it is striking that even in countries that have implemented a very light lockdown, such as Sweden, the epidemic peaked long before herd immunity was achieved. Furthermore, even societies that slowly reopened businesses and public life did not experience an immediate explosive resurgence of the epidemic expected given the low levels of immunity and the speed with which the disease spread initially [3].

Here, we will propose an agent-based lattice model of an infectious disease that spreads in a geographically heterogeneous population. The model is a simplified depiction of the dynamics of viral infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), with its characteristic parameters. We will examine the effect of the heterogeneous infection pattern that is so characteristic of this disease, using a gamma distributed infectiousness with dispersion factor $k = 0.1$ from [4]. Infection heterogeneity is a feature of several epidemic diseases [5] and plays a particularly important role for COVID-19 [4,6–8]. In individual events, a single person has caused dozens of infections [9]. At the same time, the attack rate within households has been reported to be very low, at less than 20% [10], despite prolonged close contact. This suggests that the majority of COVID-19 patients infect very little. In [11] an agent-based model was used to demonstrate that the Achilles heel of an epidemic driven by superspreaders was public social contacts, while the repeated contacts to smaller family and work groups were less dangerous.

Looking at COVID-19 data from the first wave in the United States in the analysis by [12], it is clear that a rapidly spreading epidemic occurs primarily in densely populated areas. In less densely populated areas, the epidemic onset is delayed, and in rural areas, the epidemic never really starts, with most cases appearing to be spillover from the cities. The daily per capita mortality at the first peak in [12] varies by a factor of ten between the most and least densely populated areas in the United States.

The density dependence of COVID epidemics could in principle be explained by the higher chance of meeting infected people in dense areas. However, only 1% of COVID-19 spread was outdoors in China [13], suggesting that only indoor meetings count. Furthermore, only few infections happen within households [10], suggesting that it is social visits and meetings in confined areas that facilitate infection [14]. We will base our model on these observations.

METHODS

Our model plays out on a lattice of side L with periodic boundary conditions. Each lattice site may either be empty or contain one agent. The agents can be in one of three states, susceptible, infectious or recovered. Individuals interact with their neighbours with a frequency f_{meet} which is constant and fixed along with the mean infection probability per meeting to give the desired number of secondary infections at the onset of the epidemic, R_0. The distance that agents travel we draw from the distribution $p(r) = \frac{1}{r_0} e^{-r/r_0}$ with a mean of $r_0 = 10$ sites unless stated otherwise. The real probability distribution of travel distances for cars has in Italy been determined to be an exponential function below 20 km, and then a power law with a steeper cut-off around 500 km [15]. The interaction radius r_0 is not by itself a meaningful parameter. Rather, it is the number of neighbours within this radius, given by $\pi \rho r_0^2$ (where ρ is the population density) which determines the behaviour of the system, as will be demonstrated below. A diagram of the model can be seen in Fig. 1.

COVID-19 is known to be transmitted before the onset of symptoms [16]. For simplicity, rather than using the susceptible-exposed-infectious-recovered (SEIR) modelling framework, we combined the exposed, presymptomatic state and the overlapping infectious period into one single infectious period of $1/\gamma = 10$ days. Given the geometry of the system, this leads to a measured serial interval of approximately 6 days, slightly higher than most sources estimate [17,18]. However, it is a small difference, and given the simplifications of our model we judged that it would not be meaningful to fit this parameter precisely to the observed value. Infected agents randomly leave the infectious state with a rate of γ per day, meaning that the duration of the disease is exponentially distributed. The implications of using a different distribution of disease durations are explored in the supplement.

When an infectious agent i interacts with a susceptible agent, the susceptible agent will become infected with a probability p_i that is specific to the infecting agent. We draw these probabilities from a gamma distribution with dispersion factor $k = 0.1$ [4,5], within the range observed for COVID-19 [4,6,8]. The distribution is normalized to give an average reproduction number R_0 of 3 at a population density of 1 at the onset of the epidemic.

The geographical heterogeneity of the population is modelled by placing a square 'city' of side $L/5$ on a lattice with periodic boundary conditions and lattice size L. The city has the population density 1, that is all sites are occupied, and the city population is thus $L^2/25$. The city is surrounded by 'countryside' with a population density and total population of $24\,L^2/25$.

Importantly, we assume that the rate of interactions per agent is kept the same in both city and countryside, meaning that we assume that people are equally social. If an agent in the countryside attempts to interact with an empty site, the attempt is counted as failed, and new attempts are made until the number of contacts is the same as in the city, where all sites are occupied.

Thus, people in the countryside interact with a smaller set of people while still spending the same amount of time on social activities. In a wider perspective, this proposes that density dependence of disease spreading is more due to difference in diversity of contacts than due to differences in time spent around other people. Thereby our model assumes an infection rate that depends on density, but not in a simple linear fashion as sometimes assumed [19].

We seed the disease within the city. Even if we were to seed it randomly, the city would usually be hit early on provided that the epidemic catches on.
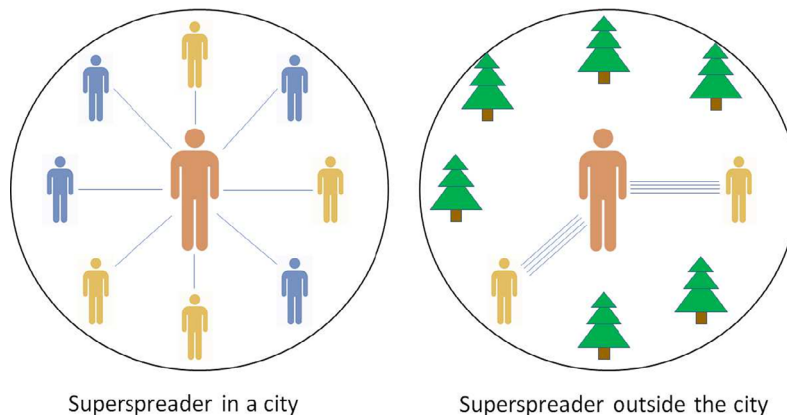

Fig. 1. Model: A superspreader in a city interacts a little with a lot of people and will infect some fraction of them. On the other hand, a superspreader outside the city will interact a lot with each of a smaller set of people. The superspreader then infects practically all of them, but there is a lower cap on the number of secondary infections.

RESULTS

In Fig. 2, we consider a disease with heterogeneous infection rates and study how simulated epidemic trajectories depend on population density. One sees that the dynamics resembles that of a SIR model at values much lower and much higher than the critical density. In the low regime, the epidemic only spreads in the city which is relatively well-mixed, whereas in the high regime, the countryside begins to resemble the city more and more. In both extremes, the fraction of infected individuals rises and falls symmetrically. The figure illustrates how the epidemic is 'stretched out' in the intermediate density range. The epidemic has the longest lifetime when is just above the percolation threshold $_{crit}$, as the disease still spreads in the countryside, but is nonetheless slowed down by the lower population density.

In Fig. 3A,B, we measure the attack rate of the epidemic in a homogeneously distributed population in order to find the percolation threshold, below which the epidemic will stop propagating. Panel (A) identifies $_{crit} \approx 0.01$ for a homogeneously spreading disease. In contrast, a disease with an overdispersion of $k = 0.1$ has a much higher critical density $_{crit} \approx 0.04$, as seen in (B). The sa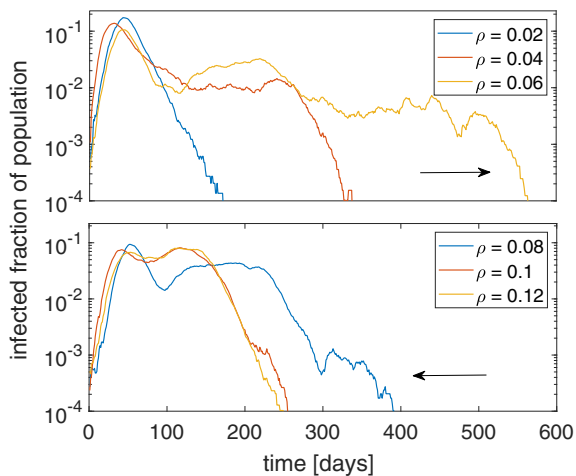me figure demonstrates that it is not the density alone that determines the ability of the disease to percolate, but rather the number of neighbours, proportional to r_0. The overdispersed simulation and the simulation with homogeneous infection were done with same average disease transmission rate and the factor ~ 4 difference in critical density comes about because a disease with $k = 0.1$ has 10% of the infected being responsible for 80% of the infections. Thus, most people do not transmit the disease, and it is therefore the density of the few people who do spread the disease that sets the critical threshold.

Since our model analysis centres on superspreaders as a main driver of the epidemic, Fig. 3C compares epidemics with and without superspreaders. It can be seen that with no superspreaders, the epidemic will spread unhindered in the countryside, albeit more slowly since the countryside is geographically larger than in the city.

This leads to a graph similar to two superimposed SIR-like models. If superspreaders are present, however, the epidemic may spread both slowly and erratically in the countryside and continue long after herd immunity is achieved in the city.

The dependence of the percolation threshold on the dispersion parameter k is further investigated in Fig. 4, which shows the attack rate as a function of both the number of neighbours and k. The figure shows that the percolation threshold increases drastically at low k, meaning that a superspreader-driven epidemic requires more social contacts per person in order to spread. Once the epidemic becomes sufficiently overdispersed ($k < 0.05$), it is no longer viable.

In Fig. 5, we try to replicate the data compiled by [12] and see that local disease incidence in a model with heterogeneous infectivity is indeed much more population density dependent than a model assuming homogeneous infectivity. This fits well with the cited data, which suggest a strong dependence of COVID-19 incidence on population density. It has already been known for years that the spread of epidemics is population density dependent [19]. Here, we show that this dependence is enhanced by heterogeneous infectivity. Importantly, as opposed to traditional disease models, we assume that everyone is equally social, but that the set of available contacts is smaller in sparsely populated regions. The significance of this will be discussed further below.

The delayed onset and erratic behaviour of the countryside epidemic obviously depend on the density and other characteristics of the countryside and the assumed travelling pattern of individuals. Also, real countryside contains a diverse pattern of smaller and large settlements. Therefore, we considered
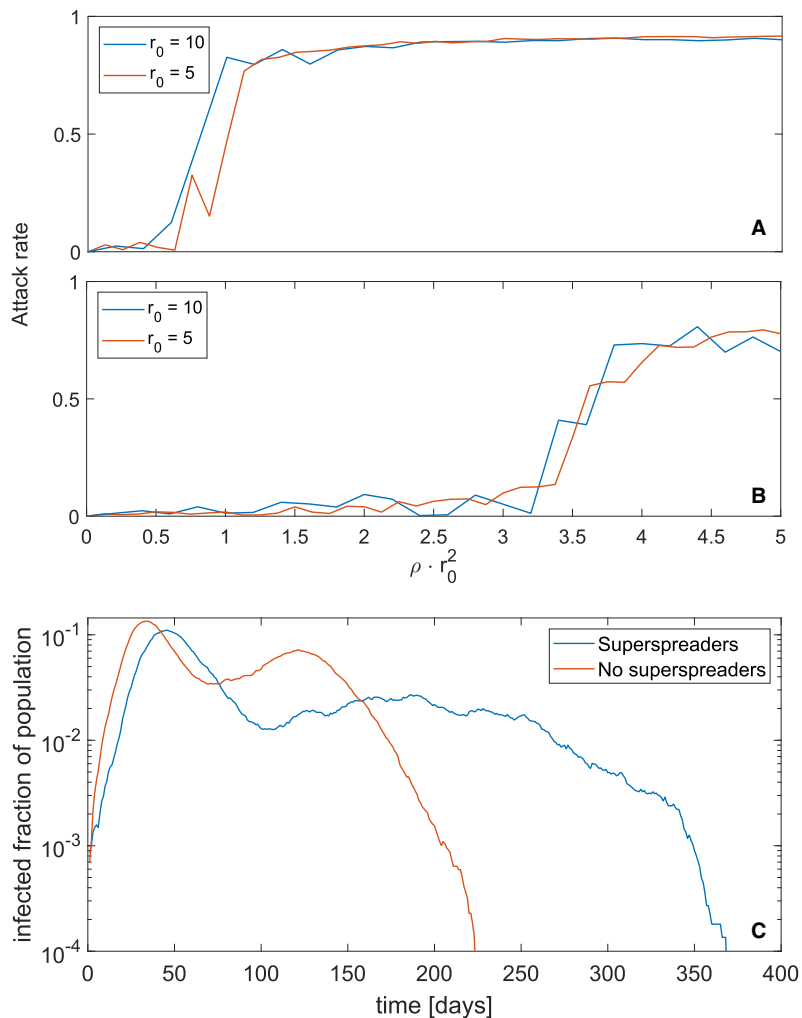

Fig. 2. Epidemic trajectories: Infected fraction of the population over time changes with the countryside population density ρ. In the low- regime, increasing the population density stretches the curve, as the epidemic spreads further from the city. When ρ is above ~ 0.06, the epidemic again approaches the behaviour of a SIR model, as the epidemic now spreads unhindered across the whole system. Around ρ_{crit}, there is a large variation in the duration of the epidemic. The parameters used are $\gamma = 0.1$, $r_0 = 10$, $f_{meet} = 10$ and dispersion parameter $k = 0.1$.

Fig. 3. Comparison of models without and with superspreaders. Panel (A) shows the attack rate as a function of the number of neighbours within the radius of interaction ($\sim \rho r_0$) in a population where everyone infects with the same rate. (B) shows attack rate with heterogeneous infection rates, using a gamma distribution with dispersion factor $k = 0.1$. The two overlaid curves demonstrate that the parameter r_0 does not affect the physics of the system, and what really determines the ability of the disease to percolate is the number of neighbours, proportional to ρr_0. (C) Epidemic trajectory when superspreaders dominate (blue) and when infectiousness is evenly distributed (red) for equal countryside population density and radius of interaction ($\rho r_0 = 6$). When superspreaders are the main drivers of the epidemic, it is strongly impeded once the city has reached herd immunity. When everyone infects equally, the epidemic simply spreads radially out from the city, leading to a 'second wave' in the countryside. Parameters are as in Fig. 2.

a system with several cities distributed on a lattice with side length $L = 1000$ (Fig. 6). The system mimics the observed city size distribution, which is fairly close to the Zipf law [20,21]. The random distribution of smaller cities is not entirely naturalistic, since evidence suggests that real cities are organized in a fractal pattern [22,23]. A figure using a city distribution closer to a fractal can be found in the supplement. For illustrative purposes, we chose a density in the countryside that is close to but below the percolation threshold for the disease ($\rho = 0.03$). With this below-critical spreading, we observe cities

that are spared and cities that are nearly completely infected. In reality assuming near critical spreading in the countryside would not be necessary for global spread, since people occasionally travel long distances [15], facilitating rare direct transmissions between distant cities.

DISCUSSION

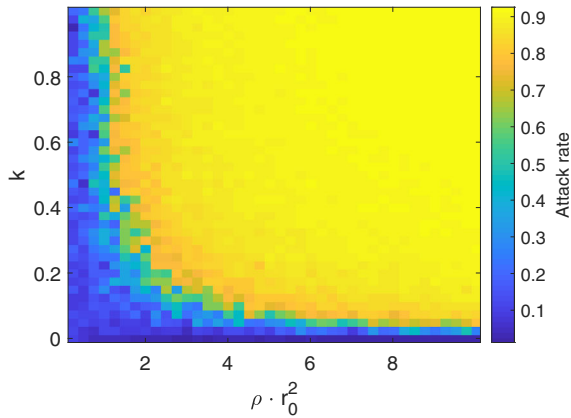Based on the above, we suggest that the lopsided appearance of COVID-19 epidemic curves can be

Fig. 4. Dependence of attack rate on density and k. Since the variable determining percolation is not the absolute density, but the number of neighbours, we plot ρr_0 on the x-axis rather than ρ. It is seen that the disease percolates much more easily at a higher k, which implies a more homogeneous infectivity. The more overdispersed the disease (corresponding to lower k), the more neighbours are required for percolation.
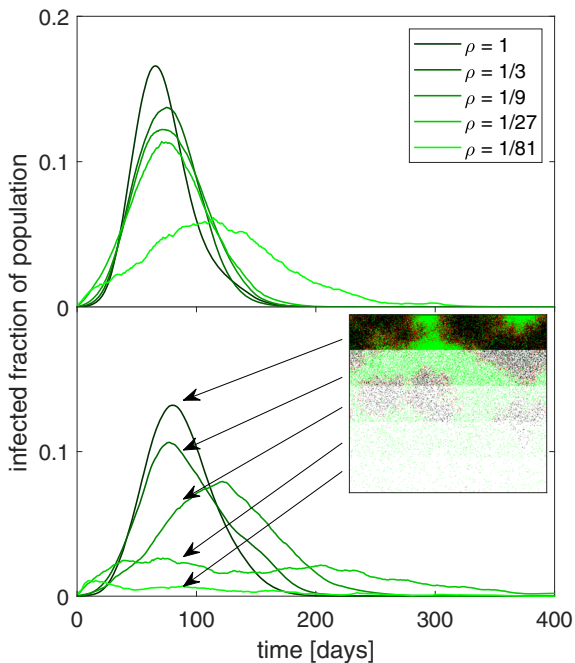

Fig. 5. Epidemic trajectory in regions of varying density. When infectivity is homogeneous (top), the epidemic is a lot less sensitive to a lower population density than when the epidemic is driven by superspreaders (bottom). Here, the epidemic is nearly absent in the low-density regions and appears to be driven by spillover. Inset illustrates the layout of the lattice.

explained by heterogeneous infection ability combined with a geographically heterogeneous population. However, our model makes a number of
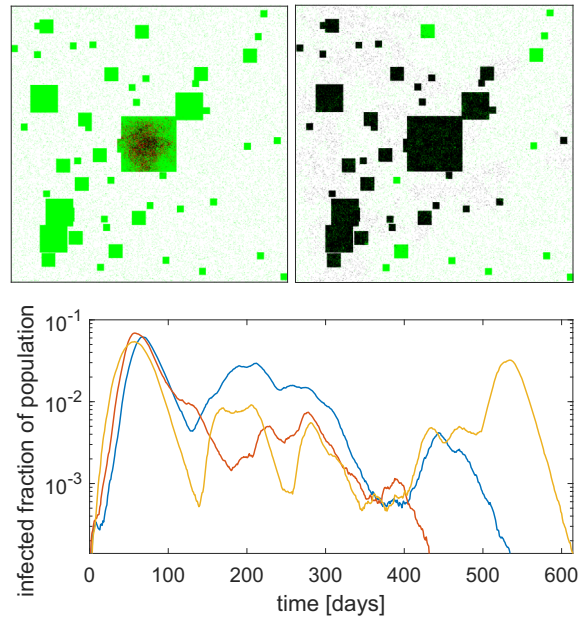

Fig. 6. Simulation with multiple cities: Infected individuals are shown in red, susceptibles are green, empty sites are white and recovered are black. City size distribution mimics Zipf's law [23], such that there is one city of with 40,000 inhabitants, 4 with 10,000, and so forth. The graphs show the fraction of the population currently infected as a function of time for three example runs of the simulation. Here, $\rho = 0.03$ and the other parameters are the same as the above figures.

assumptions and breaks from reality, whose importance must be discussed.

First and foremost, the distribution of full and empty lattice sites does not represent the geographical distribution of people as such. Rather, it represents the density of contacts. The density ratio between people in the countryside compared to an inner city is much lower than what is used in our model. On the other hand, persons living outside the cities may be more mobile than city-dwellers. The density contrast of agents on our lattice represents the combined effect of these factors.

A further complication is the distribution of disease duration and incubation periods. We here assume a simple exponential distribution of infectious period duration, whereas the real mechanics of COVID-19 includes presymptomatic transmission and a broad gamma distribution of incubation periods [17]. A different infectious period distribution might complicate our findings. Therefore, we examine the effect of a gamma distributed infectious period duration in the supplement. We find that, while changing this distribution has an effect on the percolation threshold, it does not change our fundamental conclusions.

It has long been discussed how disease transmission rate depends on population density [19,24,25]. We here present a new way of looking at this problem. A classical SIR model with pseudo-mass action transmission would assume a simple linear dependence of R_0 on population density, implicitly assuming that people become more social by living in a densely populated area.

Our model makes a different assumption: People will be equally social regardless of population density, but when population density is lower, the groups that they spend time with will be less diverse. The number of possible unique contacts for each individual declines linearly with population density, but people in sparsely populated areas are likely to have multiple encounters with the same persons, as their communities are smaller.

If, however, the epidemic is dominated by superspreaders who only need one or a few encounters to transmit the disease, the duration of each encounter becomes less important as even a rather brief contact to a superspreader is likely to lead to infection. Instead, what limits the action of superspreaders is their number of unique contacts. Superspreaders that interact with only a small, tight-knit group will inherently be highly limited in how many secondary infections they can generate.

If superspreading makes a disease vulnerable to variations in population density, we should conversely expect to see that diseases with a homogeneous infectivity, that is a dispersion factor close to or above 1, exhibit little variation with population density. One example of a disease with a homogeneous infectivity is influenza, with an estimated dispersion factor of $k = 0.94$ for the 1918 pandemic flu [26]. When examining the incidence of seasonal influenza, which we assume to have a similar dispersion factor, [27] find no consistent variation with population density. This is a point in favour of the link between superspreading and population density dependence.

Finally, the effect of lockdowns and changes of social behaviour is important. A previous paper [11] suggests that even moderate mitigation may limit the action of superspreaders by reducing the maximal number of people any person can be in contact with. If this hypothesis is true, bans on gatherings and a reduction in public social life would lead to early peaks in the number of new cases. Our study compounds this finding and suggests that a change in behaviour is not strictly necessary to cause an epidemic peak well before herd immunity has been achieved. Mitigation strategies that primarily target cities may well be sufficiently effective in bringing down the epidemic. However, large events like funerals, weddings or festivals are not included in our model and will of course facilitate spreading in any

location. Finally, if there is a large difference in attitude towards the disease in the city and the countryside, this can significantly change the behaviour of the epidemic in the countryside relative to the city. All in all, how governments and populations have responded to the pandemic has likely had a crucial effect on its trajectory. Our model investigates how it would have played out in the absence of these complicating factors.

CONCLUSION

Despite some caveats, our model reproduces the main aspects of geographical heterogeneity and suggests a new view on density dependence of epidemic dynamics. An epidemic with a large heterogeneity in infection rates is predicted to be most intense in large cities while it slowly tapers off in the countryside. This is consistent with what we see in data from the first half year of the COVID-19 pandemic [12,28]. Our results thus favour the hypothesis that the COVID-19 pandemic is driven by superspreaders, and that the observed quick exponential growth phase, early peak, and slow, erratic recovery phase are in part consequences of combining heterogeneity of infectivity with a heterogeneous population density.

Of course, restrictions limiting social contact were crucial to the overall COVID-19 pandemic, and behavioural societal changes remain a central part of a pandemic that still after a year is far from reaching herd immunity. Our model predicts that anything that limits the diversity of contacts, be it population density or limits on gatherings, will have an outsized effect on disease spread. In comparison, simply limiting the number or duration of contacts will not be as effective unless the diversity is also decreased. A superspreader still spreads quite effectively given ten encounters with ten different individuals, but is much less effective given 20 encounters with only five different people. The present work suggests that future epidemiological models should take contact diversity into account if they are to properly capture the dynamics of COVID-19 and other diseases with large transmission heterogeneity.

ETHICS STATEMENT

This study involved no experimental work, and ethical approval was therefore not relevant.

CONFLICT OF INTEREST

The authors declare no conflicting interests.

AUTHOR CONTRIBUTIONS

KS proposed the model design. AE wrote the code used for the simulations presented here, produced the figures and wrote the initial draft of the paper. KS and AE edited the paper.

DATA AVAILABILITY STATEMENT

The code used to produce the figures shown in this article is available on Figshare [29].

REFERENCES

1. Remuzzi A, Remuzzi G Covid-19 and Italy: what next? Lancet. 2020;395:1225–8.
2. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD How will country-based mitigation measures influence the course of the covid-19 epidemic? Lancet. 2020;395:931–4.
3. Dong E, Du H, Gardner L An interactive webbased dashboard to track covid-19 in real time. Lancet Infect Dis. 2020;20:533–4.
4. Endo A, Abbott S, Kucharski AJ, Funk S Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china. Wellcome Open Res. 2020;5:67.
5. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM Superspreading and the effect of individual variation on disease emergence. Nature. 2005;438:355–9.
6. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of covid-19 in 391 cases and 1286 of their close contacts in shenzhen, china: a retrospective cohort study. Lancet Infect Dis. 2020;20(8):911–9.
7. Adam D, Wu P, Wong J, Lau E, Tsang T, Cauchemez S, et al. Clustering and superspreading potential of severe acute respiratory syndrome coronavirus 2 (sars-cov-2) infections in hong kong, Research Square - preprint 2020.
8. Miller D, Martin MA, Harel N, Kustin T, Tirosh O, Meir M, et al. Full genome viral sequences inform patterns of sars-cov-2 spread into and within Israel. Nat Commun. 2020;11:5518. https://doi.org/10.1038/s41467-020-19248-0
9. Shen Y, Li C, Dong H, Wang Z, Martinez L, Sun Z, et al. Community outbreak investigation of sars-cov-2 transmissionamong bus riders in eastern China. JAMA Intern Med. 2020;180(2):1665–1671.
10. Park SY, Kim Y-M, Yi S, Lee S, Na B-J, Kim CB, et al. Early release-coronavirus disease outbreak in call center, South Korea. Emerg Infect Dis. 2020;26:1666–70.
11. Sneppen K, Taylor RJ, Simonsen L Impact of superspreaders on dissemination and mitigation of covid19. medRxiv. 2020.
12. Medhi N Covid-19 cases in the U.S. by population density. https://www.endcoronavirus.org/projects-1/covid-19-cases-in-the-us-by-population-density. Accessed 20 Aug 2020.
13. Carlson CJ, Chipperfield JD, Benito BM, Telford RJ, O'Hara RB Species distribution models are inappropriate for covid-19. Nature Ecol Evol. 2020;4:770–1.
14. Nishiura H, Oshitani H, Kobayashi T, Saito T, Sunagawa T, Matsui T, et al. Closed environments facilitate secondary transmission of coronavirus disease 2019 (covid-19). medRxiv. 2020.
15. Pappalardo L, Rinzivillo S, Qu Z, Pedreschi D, Giannotti F Understanding the patterns of car travel. Eur Phys J Spl Topics. 2013;215:61–73.
16. Tindale LC, Stockdale JE, Coombe M, Garlock ES, Lau WYV, Saraswat M, et al. Evidence for transmission of covid-19 prior to symptom onset. Elife. 2020;9:e57149.
17. Nishiura H, Linton NM, Akhmetzhanov AR Serial interval of novel coronavirus (COVID-19) infections. Int J Infect Dis. 2020;93:284–6.
18. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. N Engl J Med. 2020;382:1199–207.
19. Anderson RM, Anderson B, May RM Infectious diseases of humans: dynamics and control. Oxford: Oxford University Press; 1992.
20. Gabaix X Zipf's law for cities: an explanation. Q J Econ. 1999;114:739–67.
21. Soo KT Zipf's law for cities: a cross-country investigation. Reg Sci Urban Econ. 2005;35:239–63.
22. Yook S-H, Jeong H, Barabási A-L Modeling the internet's large-scale topology. Proc Natl Acad Sci. 2002;99:13382–6.
23. Mori T, Smith TE, Hsu W-T Common power laws for cities and spatial fractal structures. Proc Natl Acad Sci. 2020;117:6469–75.
24. Finkenstädt B, Keeling M, Grenfell B Patterns of density dependence in measles dynamics. Proc R Soc Lond B Biol Sci. 1998;265:753–62.
25. Finkenstädt B, Grenfell B Empirical determinants of measles metapopulation dynamics in England and Wales. Proc R Soc Lond B Biol Sci. 1998;265:211–20.
26. Fraser C, Cummings DA, Klinkenberg D, Burke DS, Ferguson NM Influenza transmission in households during the 1918 pandemic. Am J Epidemiol. 2011;174:505–14.
27. Bonabeau E, Toubiana L, Flahault A The geographical spread of influenza. Proc R Soc Lond B Biol Sci. 1998;265:2421–5.
28. Smith M, Yourish K, Almukhtar S, Collins K, Ivory D, Harmon A Covid-19 data repository. https://github.com/nytimes/covid-19-data. Accessed 03 Sept 2020.
29. Code available under the URL https://doi.org/10.6084/m9.figshare.12919709.v1