medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188359; this version posted September 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

COVID-19 superspreading in cities versus the countryside

Andreas Eilersen and Kim Sneppen
*Niels Bohr Institute, University of Copenhagen**

(Dated: September 4, 2020)

So far, the COVID-19 pandemic has been characterised 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 are the core factors explaining this discrepancy. We do so through 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.

INTRODUCTION

At its onset the 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 has 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 COVID-19 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 heal 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 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 peak daily per capita mortality in [12] varies by a factor of ten between the most and least densely populated areas in the USA.

MODEL

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}. 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. The real probability distribution of travel distances for cars has in Italy been determined to be an exponential function below 20 kilometers, and then a powerlaw with a steeper cutoff around 500 kilometers [13].

COVID-19 is known to be transmitted before the onset of symptoms [14]. For simplicity, rather than using the SEIR modeling framework, we combined the exposed, presymptomatic state and the overlapping infectious period (both estimated at about five days) into one single infectious period of $1/\gamma = 10$ days [15, 16].

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 normalised to give an average reproduction number R_0 of 3 at a population density of 1.

The geographical heterogeneity of the population is modeled by placing a square "city" of side $L/5$ on a lat-

* sneppen@nbi.ku.dk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188359; this version posted September 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

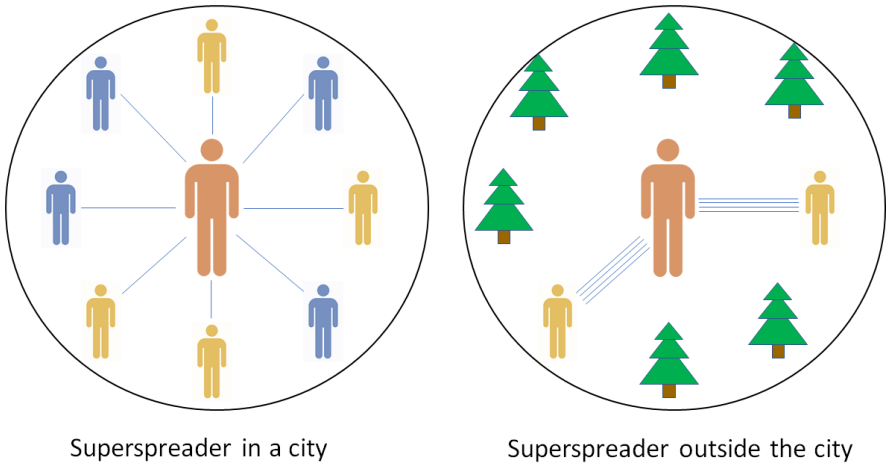

FIG. 1. **Model:** A superspreader in a city interacts a little with a lot of people and will infect some fraction of them. On the other hand, a superspreader outside the city will interact a lot with each of a smaller set of people. The superspreader then infects practically all of them, but there is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

tice with periodic boundary conditions and lattice size L. The city has the population density 1 , i.e. all sites are occupied, and the city population is thus $L^2/25$. The city is surrounded by "countryside" with a population density ρ and total population of $24\rho L^2/25$.

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 [17].

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.

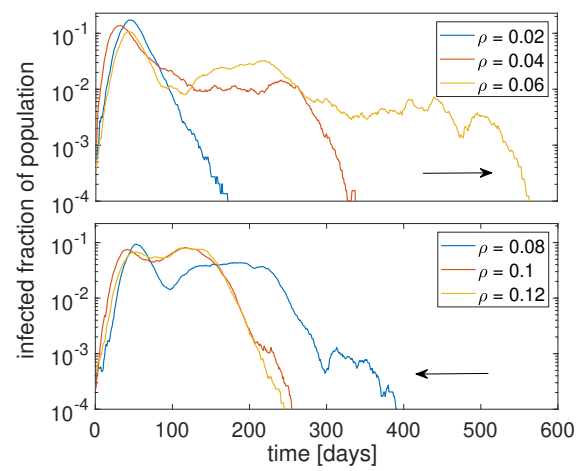

FIG. 2. **Epidemic trajectories:** Infected fraction of the population over time changes with the countryside population density ρ. In the low-ρ regime, increasing the population density stretches the curve, as the epidemic spreads further from the city. When ρ is above ~ 0.06, the epidemic again approaches 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 $k = 0.1$.

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 ρ 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 rise and fall symmetrically. The figure illustrates how the epidemic is "stretched out" in the intermediate density range. The epidemic has the longest lifetime when ρ is just above ρ_{crit}, as the disease still spreads in the countryside, but is nonetheless slowed down by the lower population density.

In fig. 3 (a, b) we measure the attack rate of the epidemic in a homogeneously distributed population in order to find the percolation threshold, below which

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188359; this version posted September 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

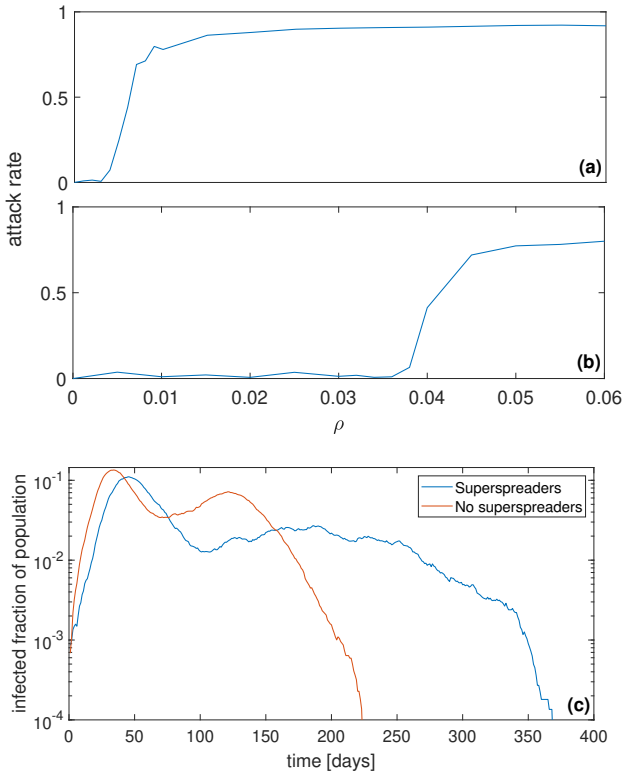

FIG. 3. **Comparison of models without and with superspreaders.** Panel (a) shows the attack rate as a function of population density in a population where all infect with the same rate. (b) shows attack rate with heterogeneous infection rates, using a gamma distribution with dispersion factor $k = 0.1$. (c) Epidemic trajectory when superspreaders dominate (blue) and when infectiousness is evenly distributed (red) for equal countryside population density. 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.

the epidemic will stop propagating. Panel (a) identifies $\rho_{crit} \approx 0.005$ for a homogeneously spreading disease. In contrast, a disease with an overdispersion of $k = 0.1$ has a much higher critical density $\rho_{crit} \approx 0.04$, as seen in (b). Notably, these simulations were done with same average disease transmission rate and the factor ~ 10 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 these superspreaders that sets the critical threshold.

Since our model analysis centers on superspreaders as a main driver of the epidemic, fig. 3(c) 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.

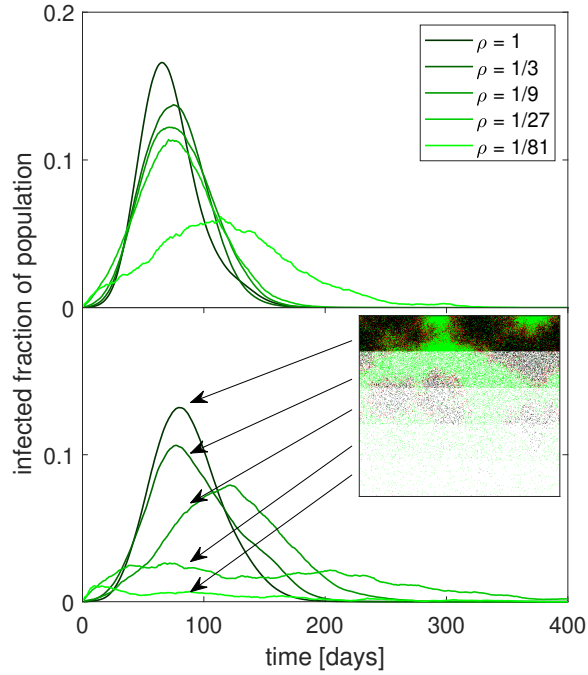

FIG. 4. **Epidemic trajectory in regions of varying density.** When infectivity is homogeneous (top), the epidemic is a lot less sensitive to a lower population density than when the epidemic is driven by superspreaders (bottom). Here, the epidemic is nearly absent in the low-density regions, and appears to be driven by spillover. Inset illustrates the layout of the lattice.

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.

In fig. 4, we try to replicate the data compiled by [12] and see that a disease 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 [17]. 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 depends 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 a system with several

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188359; this version posted September 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

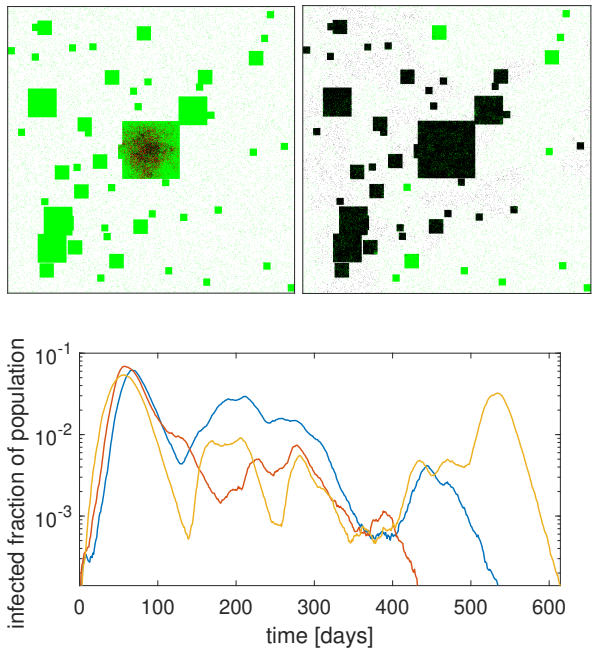

FIG. 5. **Simulation with multiple cities:** Infected individuals are shown in red, susceptibles are green, empty sites are white and recovered are black. City size distribution mimics Zipf's law [19], such that there is one city of with 40000 inhabitants, 4 with 10000, and so forth. The graphs show 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.

cities distributed on a lattice with side length $L = 1000$. The system mimics the observed city size distribution which is fairly close to the Zipf law [18]. 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 not necessary for global spread, since people occasionally travel long distances [13], 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 explained by heterogeneous infection ability combined with a geographically heterogeneous population. However, our model makes a number of 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 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.

It has long been discussed how disease transmission rate depends on population density [17, 20, 21]. 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, i.e. 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 [22]. When examining the incidence of seasonal influenza, which we assume to have a similar dispersion factor, [23] 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.

Despite some caveats, our model reproduces the main aspects of geographical heterogeneity and suggests a new

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188359; this version posted September 7, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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 COVID-19 pandemic [12, 24]. 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 recovery phase are consequences of combining heterogeneity of infectivity with a heterogeneous population density.

ACKNOWLEDGMENTS

We thank Lone Simonsen, Viggo Andreasen, and Bjarke Frost Nielsen for enlightening discussions. Our research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No. [740704].

[1] A. Remuzzi and G. Remuzzi, Covid-19 and italy: what next?, The Lancet (2020).

[2] 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).

[3] E. Dong, H. Du, and L. Gardner, An interactive web-based dashboard to track covid-19 in real time, The Lancet infectious diseases **20**, 533 (2020).

[4] A. Endo, S. Abbott, A. J. Kucharski, S. Funk, *et al.*, Estimating the overdispersion in covid-19 transmission using outbreak sizes outside china, Wellcome Open Research **5**, 67 (2020).

[5] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz, Superspreading and the effect of individual variation on disease emergence, Nature **438**, 355 (2005).

[6] Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, *et al.*, Epidemiology and transmission of covid-19 in 391 cases and 1286 of their close contacts in shenzhen, china: a retrospective cohort study, The Lancet Infectious Diseases (2020).

[7] D. Adam, P. Wu, J. Wong, E. Lau, T. Tsang, S. Cauchemez, G. Leung, and B. Cowling, Clustering and superspreading potential of severe acute respiratory syndrome coronavirus 2 (sars-cov-2) infections in hong kong, Research Square - preprint (2020).

[8] D. Miller, M. A. Martin, N. Harel, T. Kustin, O. Tirosh, M. Meir, N. Sorek, S. Gefen-Halevi, S. Amit, O. Vorontsov, *et al.*, Full genome viral sequences inform patterns of sars-cov-2 spread into and within israel, medRxiv (2020).

[9] Y. Shen, C. Li, H. Dong, Z. Wang, L. Martinez, Z. Sun, A. Handel, Z. Chen, E. Chen, M. H. Ebell, F. Wang, B. Yi, H. Wang, X. Wang, A. Wang, B. Chen, Y. Qi, L. Liang, Y. Li, F. Ling, J. Chen, and G. Xu, Community outbreak investigation of sars-cov-2 transmissionamong bus riders in eastern china, JAMA Internal Medicine (2020).

[10] S. Y. Park, Y.-M. Kim, S. Yi, S. Lee, B.-J. Na, C. B. Kim, J.-i. Kim, H. S. Kim, Y. B. Kim, Y. Park, *et al.*, Early release-coronavirus disease outbreak in call center, south korea, Emerging Infectious Diseases (2020).

[11] K. Sneppen, R. J. Taylor, and L. Simonsen, Impact of superspreaders on dissemination and mitigation of covid-19, medRxiv (2020).

[12] N. Medhi, 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` (2020), accessed: 20-08-2020.

[13] L. Pappalardo, S. Rinzivillo, Z. Qu, D. Pedreschi, and F. Giannotti, Understanding the patterns of car travel, The European Physical Journal Special Topics **215**, 61 (2013).

[14] L. C. Tindale, J. E. Stockdale, M. Coombe, E. S. Garlock, W. Y. V. Lau, M. Saraswat, L. Zhang, D. Chen, J. Wallinga, and C. Colijn, Evidence for transmission of covid-19 prior to symptom onset, Elife **9**, e57149 (2020).

[15] H. Nishiura, N. M. Linton, and A. R. Akhmetzhanov, Serial interval of novel coronavirus (covid-19) infections, International journal of infectious diseases (2020).

[16] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, *et al.*, Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia, New England Journal of Medicine (2020).

[17] R. M. Anderson, B. Anderson, and R. M. May, *Infectious diseases of humans: dynamics and control* (Oxford university press, 1992).

[18] X. Gabaix, Zipf's law for cities: an explanation, The Quarterly journal of economics **114**, 739 (1999).

[19] K. T. Soo, Zipf's law for cities: a cross-country investigation, Regional science and urban Economics **35**, 239 (2005).

[20] B. Finkenstädt, M. Keeling, and B. Grenfell, Patterns of density dependence in measles dynamics, Proceedings of the Royal Society of London. Series B: Biological Sciences **265**, 753 (1998).

[21] B. Finkenstädt and B. Grenfell, Empirical determinants of measles metapopulation dynamics in england and wales, Proceedings of the Royal Society of London. Series B: Biological Sciences **265**, 211 (1998).

[22] C. Fraser, D. A. Cummings, D. Klinkenberg, D. S. Burke, and N. M. Ferguson, Influenza transmission in households during the 1918 pandemic, American journal of epidemiology **174**, 505 (2011).

[23] E. Bonabeau, L. Toubiana, and A. Flahault, The geographical spread of influenza, Proceedings of the Royal Society of London. Series B: Biological Sciences **265**, 2421 (1998).

[24] M. Smith, K. Yourish, S. Almukhtar, K. Collins, D. Ivory, and A. Harmon, Covid-19 data repository, `https://github.com/nytimes/covid-19-data` (2020), accessed: 03-09-2020.