



Mathematical Statistics
Stockholm University

**Individual preventive social distancing
during
an epidemic may have negative
population-level outcomes**

Ka Yin Leung¹, Frank Ball², David Sirl³ and Tom Britton⁴

Research Report 2018:14

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

<http://www.math.su.se>



Individual preventive social distancing during an epidemic may have negative population-level outcomes

Ka Yin Leung* Frank Ball[†] David Sirl[‡] Tom Britton*

May 2018

Abstract

The outbreak of an infectious disease in a human population can lead to individuals responding with preventive measures in an attempt to avoid getting infected. This leads to changes in contact patterns. However, as we show in this paper, rational behaviour at the individual level, such as social distancing from infectious contacts, may not always be beneficial for the population as a whole. We use epidemic network models to demonstrate the potential negative consequences at the population level. We take into account the social structure of the population through several network models. As the epidemic evolves, susceptible individuals may distance themselves from their infectious contacts. Some individuals replace their lost social connections by seeking new ties. If social distancing occurs at high rates at the beginning of an epidemic, then this can prevent an outbreak from occurring. However, we show that moderate social distancing can worsen the disease outcome both in the initial phase of an outbreak and the final epidemic size. Moreover, the same negative effect can arise in real-world networks. Our results suggest that one needs to be careful when targeting behavioural changes as they could potentially worsen the epidemic outcome. Furthermore, network structure crucially influences the way that individual-level measures impact the epidemic at the population level. These findings highlight the importance of careful analysis of preventive measures in epidemic models.

*Stockholm University, Sweden; kayin.leung@math.su.se

[†]University of Nottingham, UK; frank.ball@nottingham.ac.uk

[‡]University of Nottingham, UK; david.sirl@nottingham.ac.uk

[§]Stockholm University, Sweden; tom.britton@math.su.se

1. Introduction

Mathematical models for the spread of infections have been successfully used to increase understanding of how epidemics may propagate: what are the most important features to determine the initial epidemic growth, final epidemic size or endemic level? Mathematical models are also useful to evaluate the possible effects on epidemic dynamics of preventive measures. This can guide public health officials to decide what measures could be put in place to reduce or even stop spreading of a disease [1].

To prevent or control an epidemic, public health authorities may implement measures by e.g. isolating/treating detected infectious cases or starting a vaccination scheme, either before or during the outbreak [1]. In addition, individuals may take their own measures to prevent themselves from getting infected, e.g. by wearing face masks, taking hygienic measures such as hand washing, or by socially distancing themselves from infectious contacts. Such individual behaviour has been observed in e.g. the recent Ebola outbreak and the 2009 A/H1N1 epidemic [2-6].

In general, it is hard to predict the effect of preventive measures without using models to guide us. Epidemic dynamics are highly nonlinear and therefore preventive measures can lead to counter-intuitive effects. Standard epidemic models assume human behaviour is not influenced by the epidemic and is constant over time. Although it is often recognized that humans do take preventive measures in the course of an epidemic, models that incorporate behavioural dynamics are generally much harder to analyze. Recently, such models have started to receive more attention, and important advances have been made to gain understanding of the effect of different behavioural changes on epidemic dynamics [7-10].

A crucial modelling ingredient is the contact pattern in the population as infection is transmitted through contacts between susceptible and infectious individuals. Owing to challenges in their analysis, the majority of models that consider behavioural responses to epidemic dynamics are relatively simple in modelling contact patterns [10]. Often the simplest assumption of homogeneous mixing, or some variant, is made. This assumption implies that any two individuals rarely meet more than once in a large population. To overcome the restriction of the lack of repeated contacts, network epidemic models have been proposed to model human contact patterns. This class of models have received much attention over the last 20 years or so [11,12]. In these models, individuals are socially connected in the network and infection is only possible along connections. Network models are also a natural way to incorporate heterogeneity in the number of connections that individuals in the population have. Throughout this paper, we refer to two individuals that are connected to each other as 'neighbours'. Exactly what a neighbour is depends on the social structure under consideration, e.g. one may think of the neighbours as 'colleagues' in workplaces or 'sexual partners' in sexual networks.

In the current paper we study a network SIR epidemic with preventive social distancing. We consider the setting where susceptible individuals distance themselves from their neighbours who they find out are infectious, perhaps sometimes simply dropping such connections and other times, in their wish to maintain a certain number of social connections, by seeking new connections (which we refer to as 'rewiring').

We study the impact of social distancing on model networks as well as real-world networks.

We show that rational preventive individual-level behaviour can have counter-intuitive negative population-level consequences. From the perspective of an individual who distances him/herself from an infectious individual, this preventive behaviour is always rational in the sense that it decreases the risk of him/her getting infected during the epidemic outbreak (here 'always' means for all rewiring and dropping rates on all networks). If the social distancing occurs at a high enough rate at the beginning of an epidemic, then this can prevent an outbreak from occurring. In such cases, the population-level effect is obviously always positive. However, we also show that having individuals who rewire away from infectious neighbours and possibly replace them with new ties *may be harmful* for the community as a whole. Depending on the network structure of the population, social distancing may in fact *increase* the epidemic threshold parameter from below to above its threshold value, making a large outbreak possible where without social distancing it was not. We also show that social distancing can *increase the final size* of the epidemic. It is important to stress that these features do not hold for all networks. However, we show that there are real-world networks as well as model networks which exhibit these properties. It is difficult to characterize completely when such individual preventive behaviour is harmful, but it tends to happen more easily if: a) the epidemic threshold parameter for an epidemic to take off (for the baseline setting without social distancing) is large, b) the network has many individuals with low degree and possibly other groups being highly inter-connected, and c) connections are more likely to be rewired than dropped.

2 Model

2.1 SIR epidemic with social distancing on a network

We consider a population in which individuals are socially connected. Two individuals that are connected to each other are referred to as neighbours and contacts are only made between neighbours. The individuals and the connections between them together make up the network structure of the population. The stochastic SIR (susceptible-infectious-recovered) epidemic with social distancing on a network is as follows. Initially, usually one individual is infectious, we call this individual the index case, and all others in the population are susceptible (specific assumptions concerning the index case are given later). An individual that gets infected becomes infectious and remains so for an exponentially distributed time with mean $1/\gamma$. During its infectious period an individual transmits infection at a constant rate β independently to each susceptible neighbour. Moreover, a susceptible individual that has an infectious neighbour distances him/herself from this neighbour. The susceptible individual then either rewires the connection to an individual chosen uniformly at random from the population or drops the connection completely. We model this by a social distancing rate ω and a probability α to rewire rather than drop the connection. Whenever a social distancing event happens, the susceptible individual immediately chooses a new neighbour uniformly at random from the entire population with probability α , and with the remaining probability $1-\alpha$ the susceptible individual simply

drops the connection (so a susceptible individual rewires from an infectious neighbour at rate $\alpha\omega$ and drops the connection at rate $(1 - \alpha)\omega$). Dropping and rewiring events happen independently between all pairs of susceptible and infectious individuals. The epidemic continues until there is no connected susceptible-infectious pair of individuals.

Note that the preventive measure of social distancing is always beneficial from the individual perspective. Indeed, a susceptible individual that distances itself from an infectious neighbour avoids the risk of getting infected by that particular individual. In the case that it chooses to replace that social connection (rewiring), and that new neighbour is recovered (and immune), transmission can no longer occur through that connection. If the neighbour is susceptible, transmission through that connection could occur later on in the epidemic. If the neighbour is infectious, then all that has happened from an epidemic point of view is that one infectious neighbour is replaced by another one, and the risk of becoming infected is unchanged. Obviously, the most beneficial option from the point of view of avoiding getting infected is not to replace the connection (corresponding to $\alpha = 0$ and $\omega > 0$ in the model). At the population level this means that there are less connections through which the epidemic can spread. Therefore, this extreme case of dropping connections is always beneficial from both the individual and population perspective (in fact, one can mathematically analyse the model with only dropping of edges (F Ball, T Britton, KY Leung, D Sirl (2018). An SIR network epidemic model with preventive dropping of edges. *Manuscript in preparation*)). Consequently, provided infectives can recover ($\gamma > 0$), if most of the social distancing is done through dropping connections rather than rewiring them (small α) then this will also be beneficial for the population.

The epidemic with social distancing is studied on two network models as well as two real-world networks. The networks are described in Section 2.2 below. Our results in Section 3 involve several epidemiological measures for the beginning and the end of the epidemic, these concepts are introduced in Section 2.3.

2.2 The networks

2.2.1 Configuration network

The configuration model is a well-studied network, both within and without the context of epidemic models [13-15]. The network is constructed by first defining its degree distribution $\{p_d\}$, where p_d is the probability that an individual has exactly d connections. In a population of size n , each of the n individuals picks a degree independently from $\{p_d\}$ and attaches that many half-edges to itself. Half-edges are then paired completely at random and the corresponding individuals are connected in the network. By way of this construction, some imperfections may arise, such as self loops or multiple connections between some pairs of individuals. However, such imperfections become sparse in the network as the population size $n \rightarrow \infty$ if the degree distribution has finite variance (see e.g. [16, Theorem 3.1.2] and *SI Section S5*). Under such conditions the asymptotic $n \rightarrow \infty$ results in our paper hold also if the network is conditioned to have no such imperfections (see [17]). Those asymptotic results are valid as approximations only for large populations. What constitute large depends on many factors but simulations indicate that usually the approximations are good for sizes in the low hundreds.

2.2.2 Clique network

The clique-network model [18] (also referred to as household-network model when the unit under consideration is interpreted as a household) has two types of connections: global network connections and clique connections. The global network structure is obtained through the configuration network with prescribed degree distribution $\{p_d\}$. On top of this, the community is partitioned into distinct units (cliques) of size three (see *SI Section S2* for a discussion on allowing for various clique sizes). The population can be partitioned into cliques by labelling all individuals from 1 to n , and letting the first three individuals make up clique 1, the next three individuals make up clique 2, and so on. In the final network, individual 1 is then connected to all individuals he/she is connected to from the construction of the configuration model together with individuals 2 and 3 from the clique construction, and similarly for the other individuals. As with the configuration network, the clique configuration network can be treated as a simple undirected network.

2.2.3 Real-world networks

The real-world networks for our studies are taken from the Stanford large network dataset collection [19], where datasets for several different networks are freely available. We considered the 'arXiv General Relativity collaboration network' and the 'Facebook social circles network'. Both networks are undirected. The arXiv General Relativity collaboration network describes scientific collaborations between authors that submitted papers to the arXiv in the General Relativity and Quantum Cosmology category. Edges between nodes represent two co-authors that have written a paper together. In the Facebook social circles network, nodes are survey participants of the social network website Facebook that were using a specific app. Edges between nodes represent the 'circles' or 'friends lists' of those participants. The networks are described in more detail using summary statistics such as degree mean, median and variance, numbers of nodes and edges in *SI Section S3.1*. Moreover, additional summary statistics such as clustering coefficients are documented in [19].

2.3 Epidemiological quantities: threshold parameters, the probability of a major outbreak, and final size

In general, the social distancing model is challenging to analyze mathematically (see [20] for analysis of the beginning of an epidemic on the configuration network). As the network structure depends on the epidemic dynamics, models very soon become intractable. Therefore, in the main text we present the heuristics of our analytical results and refer to *SI* for the mathematical details. In Section 3 the main focus is on our findings from simulation studies. Here, we present the key epidemiological concepts that are used in Section 3.

For the beginning of the epidemic, in the configuration network model we use the basic reproduction number R_0 that has the interpretation as the expected number of secondary cases generated by one typical newly infected individual at the beginning of the epidemic. The number R_0 is a threshold parameter with threshold value one in the sense that, in the limit as the population size $n \rightarrow \infty$ there is a positive probability of a major outbreak (one which infects a strictly positive fraction of the population as $n \rightarrow \infty$ if $R_0 > 1$ and no major outbreak occurs if $R_0 \leq 1$). Owing to stochastic effects, it is always possible that an epidemic dies out when introduced into a population (with finite size n even when $R_0 > 1$). Previous work ([20]; see *SI Section*

SI.2) showed that the basic reproduction number R_0 for the epidemic on the configuration network with social distancing is given by

$$R_0 = \frac{\beta}{\beta + \omega + \gamma} \left(\mu_D + \frac{\sigma_D^2}{\mu_D} - 1 \right)$$

where μ_D and σ_D^2 are the mean and variance of the degree distribution $\{p_d\}$ of the configuration network, $\mu_D + \frac{\sigma_D^2}{\mu_D} - 1$ is the expected number of susceptible connections of a typical newly infected individual in the early stages of an epidemic and $\beta/(\beta + \omega + \gamma)$ is the probability of transmitting to such a susceptible individual before he/she recovers or the neighbour drops the connection or rewires away.

Related to R_0 is the clique reproduction number R_* (also referred to as the household reproduction number when the cliques under consideration are households), which is more natural to consider when studying populations with a clique structure. Rather than considering a newly infected individual, one considers a newly infected clique as the unit of interest. The same threshold behaviour holds. The clique reproduction number R_* is derived in Section 3.2 and *SI Section S2.1*.

For an epidemic on both the configuration network and the clique network, as population size n tends to infinity, the final fraction \bar{Z}_n of individuals that ever get infected converges in distribution to random variable \bar{Z} with two-point distribution: $P(\bar{Z} = 0) = 1 - P(\bar{Z} = z)$. In the event of a major outbreak, the limiting final fraction of the population infected by the epidemic is z . In general, this constant z is only characterized implicitly, even for the simplest Markovian homogeneously mixing SIR epidemic model. We use the practical definition in our simulation studies in Section 3 that an epidemic outbreak is major if the final number of infected individuals is more than 10% of the total population size. We use the fraction of simulations resulting in major outbreaks according to this definition as an approximation for the probability of a major outbreak to occur. Furthermore, we set the mean infectious period $1/\gamma$ equal to 5 days as this is a typical for many infectious diseases. More details on the simulation studies are provided in *SI Section S5*. We call the model without social distancing ($\omega = 0$) the baseline model.

3 Results

3.1 The configuration network

Social distancing in the configuration network is always beneficial at the beginning of an epidemic in the sense that it lowers R_0 . This conclusion follows immediately from expression (1). In fact, social distancing can ensure that R_0 is reduced below the epidemic threshold value of one, see Fig. 1B for an example. At the beginning of an epidemic, from the point of view of a susceptible individual, social distancing from an infective neighbour ensures with high probability that he/she avoids infection during the early stages of an epidemic. Indeed, there are only few infectives in the population in that stage of the epidemic. This makes it unlikely for a susceptible individual to encounter another infectious individual at the beginning of the epidemic.

However, social distancing need not be beneficial for the population as a whole. In fact, even though rewiring decreases R_0 , it can still lead to an increase in the final

size. To show analytically that the expected final size can increase with ω we consider a very specific degree distribution, where individuals have either degree 0 or degree k , where $k > 2$, i.e. $p_0 = 1 - p_k$ (proving things for more general degree distributions seems very hard). We analyze a related model that allows us to derive an asymptotic lower bound for the model of interest with strictly positive rewiring probability $\alpha > 0$. In the related model, we consider an SI infection ($\gamma = 0$). Then continuity arguments ensure that our results also hold for an SIR infection with $\gamma > 0$ small enough. Individuals act differently depending on their degree. A susceptible individual that tries to rewire to a randomly chosen individual v in the population will not do so if v is of degree k . If v is of degree 0, then rewiring takes place as usual, but v is prohibited from transmitting to other individuals. Therefore, the number of infections in the modified model is always less than in the original model (and is equal in the baseline model when there is no social distancing). For this modified model, we can derive an asymptotic (as $n \rightarrow \infty$) lower bound for the final size that is increasing in ω for small $\omega > 0$. It follows that, provided $\alpha > 0$, for sufficiently small $\gamma > 0$, the final size of the model with social distancing is greater than that without social distancing for sufficiently small $\omega > 0$. The details of the analysis are found in *SI Section S1.3*.

Rather than providing details for the analytical results for the final size here, we demonstrate the negative population level effects through simulation studies. We consider the social distancing model on a configuration network with heterogeneous degree distribution in Fig. 1. Parameter values are such that the basic reproduction number R_0 is large in the baseline setting and the majority of the social distancing is done through rewiring rather than dropping. The epidemic is started with 10 index cases (chosen uniformly at random from the population) in order to have most of the simulations resulting in major outbreaks. The number of index cases, unless sufficiently large, does not affect the final size of a major outbreak. We illustrate this fact by considering the scenario with one index case in *SI Fig S1*. Then the final size given a major outbreak increases as a function of social distancing as in Fig. 1, but the fraction of simulations resulting in a major outbreak is much smaller and consequently the average final size is decreasing. Additional results showing that social distancing can increase the final size for several other configuration network models are presented in *SI Section S4*. In particular, we consider different settings with a smaller rewiring probability $\alpha > 0$.

Note that the fraction of epidemics that result in major outbreaks decreases with increasing social distancing rates (Fig. 1B). Despite this, the average final size of all outbreaks can still increase. Once the social distancing rate ω increases to a level such that the basic reproduction number drops below the epidemic threshold value of one (Fig. 1B), mostly minor outbreaks will occur. Finally, we note that deviations from the average final size are generally small (also compared to the total population size of 5000), especially when conditioning on the occurrence of a major outbreak.

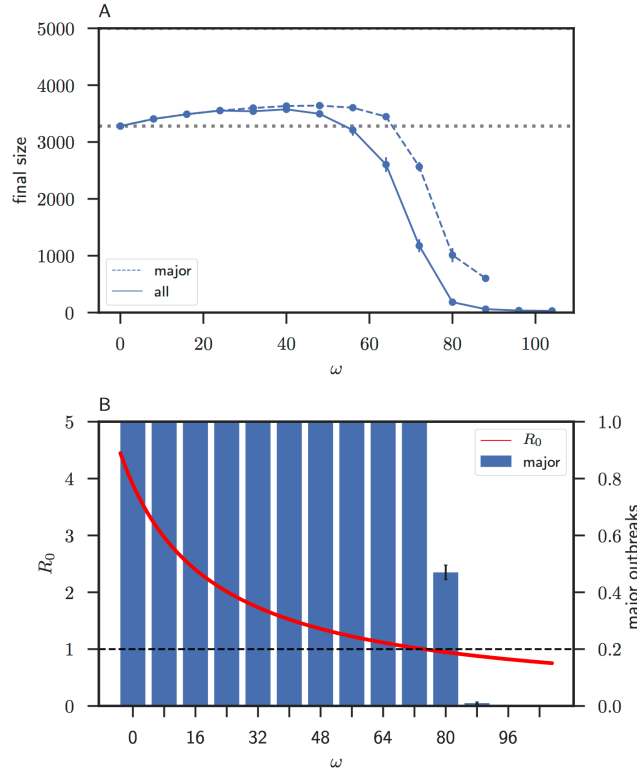


Figure 1: Social distancing can lead to an increase in the final size for the configuration network model. (A) Average final size (with 95% confidence intervals (CI) whenever large enough to be visible on the scales used in the plots) over all outbreaks (solid line) and restricted to major outbreaks (dashed line); the dotted horizontal line is at the final size when $\omega = 0$, for reference. (B) R_0 as a function of social distancing rate ω (dashed black line at $R_0 = 1$ indicates the threshold value) and fraction of all outbreaks resulting in major outbreaks (with 95% CI). Model parameters are as follows. An individual in the population has degree d with $d = 0, \dots, 10$ with probability $p_d = c/(d + 1)$, $d = 0, \dots, 10$ with $c = 0.331$ the normalization constant. Other parameter values are $\alpha = 0.9$, $\beta = 200/\text{day}$ and $1/\gamma = 5$ days, total population size 5000, and each epidemic starts with 10 randomly chosen index cases. For each value of ω , 500 epidemics are simulated.

3.2 The clique network

In the clique network individual preventive social distancing can have a negative population-level effect already at the beginning of an epidemic. To demonstrate this we consider R_* for the clique-network model. The clique reproduction number R_* is derived by differentiating between two types of newly infected cliques. A newly infected clique at first consists of one newly infected individual while the remaining clique members are susceptible. The two types are determined by the way the newly infected individual u_* was infected: (1) u_* was infected by a global neighbour (i.e. outside his/her own clique) that it had already before the start of the epidemic or (2) u_* was infected by a global neighbour that it acquired through a social distancing event during the epidemic. The clique reproduction number is the dominant eigenvalue of the 2×2 matrix $(K_{ij})_{i,j=1,2}$, where K_{ij} is the expected number of cliques of type j generated by one newly infected clique of type i . Details of the derivation of the K_{ij} are found in *SI Section S2.1*. We find an explicit expression for R_* that we can analyse as a function of social distancing ω and rewiring probability α for different degree distributions (see *SI Section S2.2*). We illustrate these analytical results with numerical examples in Fig. 2 for fixed rewiring probability $\alpha = 0.9$ (but

note that there is generally a larger range for α for which negative population effects can occur depending on the network under consideration and other model parameter values, see *SI Section S4*).

As can be seen in Fig. 2A, R_* can increase as a function of the social distancing rate ω . In particular, social distancing can move the epidemic threshold R_* from below to above its threshold value of one. In other words, individual preventive measures that are beneficial at the individual level can cause a major outbreak to become possible while without the preventive measures this is not possible. However, this depends heavily on the precise network structure. In Fig. 2B, the degree distribution is chosen such that R_* decreases for all social distancing rates. See *SI Section S2* for more details and examples of the dependence of R_* on social distancing. Note that R_* will eventually decrease for large enough social distancing rates as can be seen in Fig. 2A.

In settings where social distancing pushes R_* from below to above the threshold for an epidemic to occur, the effect of social distancing on the final size is large (Fig. 2C). Moreover, even in settings where social distancing reduces R_* , the final size can initially increase when social distancing is introduced into the model (Fig. 2D).

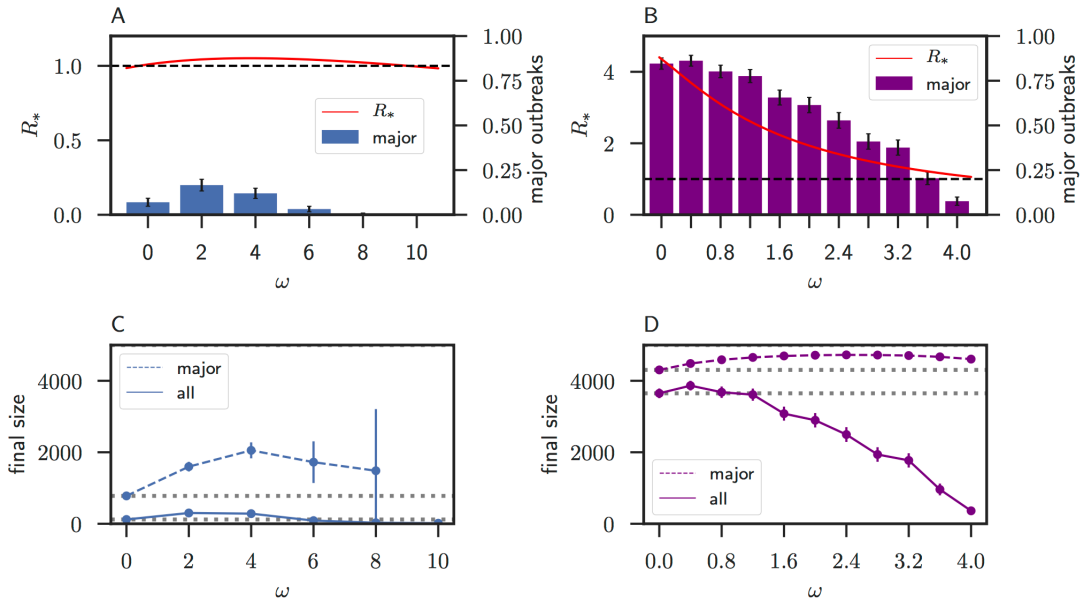


Figure 2: The effect of social distancing on the epidemic threshold parameter R_* and the final size. The fraction of epidemics resulting in major outbreaks (with 95% CI whenever large enough to be visible on the scales used in the plots) and R_* for (A) mean infectious period $1/\gamma = 5$ days, $\beta = 20/\text{day}$ and two-point degree distribution with $p_0 = \frac{1}{2} = p_1$ and (B) mean infectious period $1/\gamma = 5$ days, $\beta = 2/\text{day}$ and two-point degree distribution with $p_0 = \frac{1}{2} = p_3$. Average final size with (dashed) and without (solid) conditioning on a major outbreak (with 95% CI) corresponding to (C) scenario A (D) scenario B; dotted horizontal lines are for comparison with the size at $\omega = 0$. Other parameter values are as follows: cliques have size 3, the population size is 5000 and $\alpha = 0.9$. Each epidemic is initiated with one randomly chosen infected individual and for each value of ω , 500 epidemics are simulated.

3.3 Application to real-world networks

We consider two real-world networks: the arXiv General Relativity collaboration network and Facebook social circles network, taken from [19]. We simulate SIR epidemics with social distancing on these two real-world networks (see *SI Section*

S3.1 for details). In Fig. 3 we demonstrate that social distancing can have a negative effect at the population level by increasing the final size in the collaboration network.

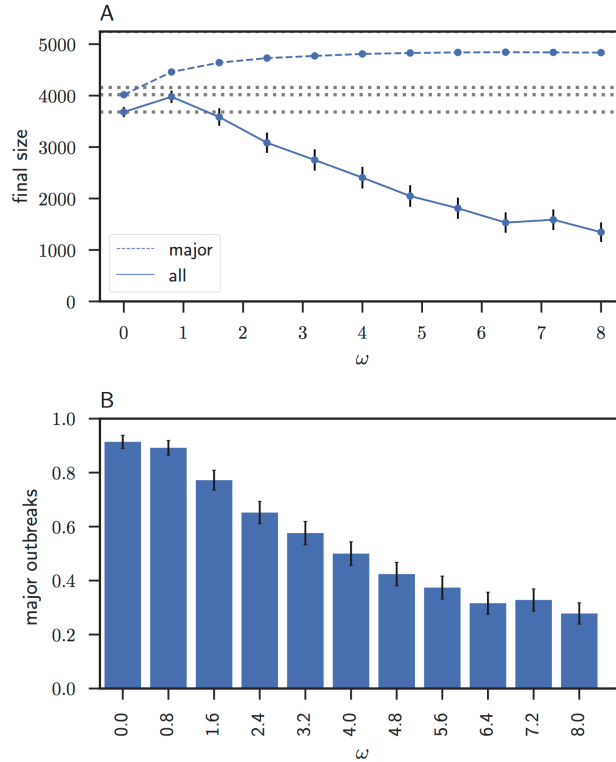


Figure 3: Social distancing can increase the final size of the epidemic on real-world networks for large recovery rate. Social distancing in the arXiv General Relativity collaboration network. (A) The average final size with (dashed) and without (solid) conditioning on a major outbreak (with 95% CI whenever large enough to be visible on the scales used in the plots); dotted horizontal lines are for the size of the giant component (top) and comparison with the size at $\omega=0$ (bottom two). (B) Fraction of all outbreaks that resulted in major outbreaks (with 95%). Model parameter values are: mean infectious period $1/\gamma = 5$ days, $\beta = 2/\text{day}$ and $\alpha = 0.9$. For each value of ω , 500 epidemics are simulated. The index case is chosen uniformly at random from the sub-population of individuals that has median degree and are part of the largest connected component of the network.

The second real-life network that we consider the social distancing epidemic model on is the Facebook social circles in Fig. 4. This serves to demonstrate that the precise network structure plays a crucial role for the effect that social distancing can have on the final size. We find that if we restrict to only the major outbreaks, then a modest increase in the final size can be observed when compared to the baseline setting. On the other hand, the average final size is more or less unaffected by social distancing for sufficiently small social distancing rates. This can be explained by the network structure of the underlying population. Since all individuals are part of the same connected component that contains many connections, i.e. all individuals are (indirectly) connected to each other, modest social distancing rates will not change the network structure in a way that significantly alters transmission patterns (see *SI Section S3.1* for network summary statistics).

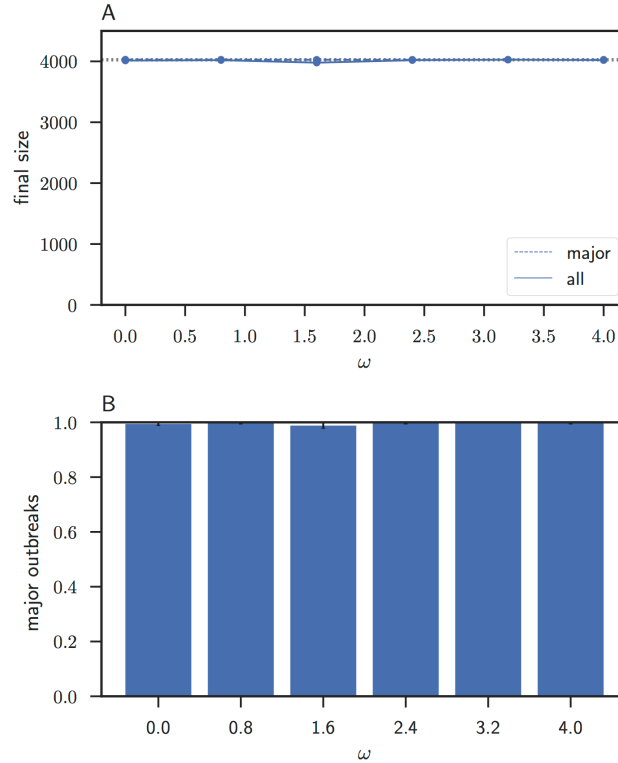


Figure 4: Social distancing in the facebook social circles network with randomly chosen index case with median degree. (A) The average final size over all outbreaks (solid) and conditioning on major outbreaks (dashed) (with 95% CI whenever large enough to be visible on the scales used in the plots); dotted horizontal lines are for the size of the network (top) and comparison with the size at $\omega=0$ (bottom two). (B) Fraction of all outbreaks that resulted in major outbreaks (with 95% CI). Model parameter values are: mean infectious period $1/\gamma = 5$ days, $\beta = 2/\text{day}$ and $\alpha = 0.9$. For each value of ω , 500 epidemics are simulated. The index case is randomly chosen from the population that has median degree.

In *SI Section 4* additional scenarios for smaller probabilities α for both real-world networks are considered. We find that negative population-level effects can occur for arXiv General Relativity collaboration network for a wider range of α -values while α has minor effects on the final size on the Facebook social circles network.

4. Conclusion and discussion

In the event of an epidemic outbreak in a population, individuals may take preventive measures by changing their contact patterns. Individuals may try to avoid infection by social distancing from infectious contacts. If this is done at sufficiently high social distancing rate, then this can have a positive population level effect by bringing the reproduction number for an epidemic to take off below the threshold value of one. On the other hand, while preventive social distancing, also at moderate rates, is always rational at the individual level, it may be harmful at the population level. In fact, preventive social distancing can increase the final epidemic size at the population level and thus have negative effects for the community at large. We demonstrated this counter-intuitive result by means of different epidemic network models, as well as simulating epidemics with social distancing on existing real-world networks. Similar conclusions in terms of behavioural changes at the individual level and its population-

level consequences have been drawn in [21,22] for different behavioural change models. Both [21,22] considered changes in human mobility patterns in the event of an epidemic and its consequences for the geographical spread. Using a metapopulation model, they illustrated that individual preventive measures in mobility patterns can lead to epidemic spread in new locations, although their invasion thresholds are always increasing [21] or even independent [22] of the behavioural changes, which is quite different from the dependence on social distancing of the threshold parameters R_0 and R_* in our models.

Whether or not social distancing of susceptible individuals from their infectious contacts will actually have negative epidemic outcomes depends strongly on the social network structure of the population. We demonstrated that social distancing can have different effects in the initial stages of the epidemic compared to the overall epidemic outbreak size. We considered the spread of an SIR epidemic on the clique-network model and the configuration network model. We showed that social distancing can have negative effects for the community by (i) increasing the epidemic threshold parameter R_* from below to above the threshold value of one in clique-networks with high clustering and (ii) by increasing the final size. Point (ii) for the final size was shown in (a) configuration networks with heterogeneous degree distribution, (b) clique-networks, and (c) a real-life collaboration network.

In general, in the baseline setting that an epidemic outbreak may occur when no preventive measures are taken, social distancing can always have beneficial effects provided that the rate of social distancing is sufficiently large (e.g. Fig. 3A). Indeed, sufficiently large social distancing rates can prevent an epidemic from taking off by reducing the epidemic threshold parameter from above to below its threshold value. In such cases, social distancing ensures that only a small number of individuals get infected by the epidemic, while in the baseline setting a significant fraction of the population may be infected.

Whereas social distancing never increases one's own risk of getting infected in our model, through rewiring, it can increase the risk for other individuals, e.g. by connecting to individuals that were previously not (so heavily) exposed to the epidemic. How and whether or not social distancing affects the population-level epidemic outcome depend on a variety of factors. Most notably, the network structure plays an important role (e.g. Fig. 2). While it was not our aim to investigate models for specific diseases, we have chosen parameter values (e.g. $1/\gamma = 5$ days and $R_0 = 4.5$ in the baseline model for Fig. 1) which are relevant for many infectious diseases. For example mumps, rubella, and polio have estimated basic reproduction numbers around 5 and infectious periods are typically in the range of a few days to weeks. Note that we consider an infectious period that is exponentially distributed. The memoryless property of the exponential distribution ensures that social distancing is always beneficial at the individual level. Relaxing this assumption could potentially lead to different effects for the individual and/or the population level. This is an interesting extension to investigate in future work for which the current framework provides an excellent starting point. Furthermore, social distancing with larger values of α can more easily lead to negative effects at the population level. Exactly what constitutes sufficiently large α to realise this effect depends on the precise setting that one considers (e.g. for the clique network model with a Poisson degree distribution this occurs with R_* for all $\frac{1}{2} < \alpha \leq 1$ while for the same model with degree

distribution $p_0 = \frac{1}{2} = p_1$ the range of α is larger at $\frac{1}{3} < \alpha \leq 1$; see also *SI Section S1.3, S2.2 and S4*). The same seems to apply when R_0 or R_* is high and the community has many individuals with low degrees and/or the community has highly connected cliques. In such cases, rewiring may introduce or increase connections to otherwise relatively isolated individuals. In this way the smaller chance of the individual who takes preventive measures getting infected is outweighed by the increased risk of transmission to a larger part of the population in the event of infection.

The main point in the paper was to show, mathematically in the supplementary material and by means of simulations in the main text, that social distancing may for some networks actually increase the total number of infected at the end of the outbreak. Social distancing could also affect other features of an outbreak, such as the size and time of the peak and the duration of the outbreak. To show any mathematical results for such finer details of the outbreak appears to be very hard but can of course be addressed by means of simulations. A thorough study, preferably accompanied by some mathematical results, remains to be done.

Although it is generally recognised that individual preventive measures are often taken once awareness of an epidemic is in place, it is not well understood how to model changes in individual behaviour. Here we considered the effect of social distancing on an epidemic. We modelled this on a contact network by assuming that susceptible individuals distance themselves from infectious contacts, allowing for both dropping of connections and replacement with new contacts in the desire to sustain a certain number of social contacts. Social behaviour is far more complex than our social distancing model, and many behavioural changes will depend on the epidemic and population under consideration. An important factor is e.g. risk perception. In the case of severe diseases, one can imagine that susceptible individuals will more likely drop connections rather than rewire them to other individuals in the population. There might be heterogeneity in preventive measures taken; some individuals might be willing to take more risks than others or have a stronger inclination to maintain a certain number of connections, e.g. for sexually transmitted infections (STI) one can often distinguish between groups with distinctively different levels of sexual activity. How such structures influence epidemic outcomes will likely depend strongly on assumptions made on e.g. mixing between risk groups (how assortative mixing is and whether individuals have the same assortative behaviour when rewiring to other individuals).

In terms of different types of connections, another interesting extension is to distinguish between behavioural changes within and between cliques. If cliques represent e.g. households then one can imagine that susceptible individuals may drop connections to infectious individuals outside the household and instead intensify connections within the household instead. While the current study focuses on preventive behavioural changes of *susceptible* individuals, one could also consider behavioural changes of infectious individuals, e.g. isolation, either self imposed or implemented by public health authorities. Such measures regarding infectious individuals would generally not have the negative population effects as seen with social distancing of susceptible individuals, though see [23] which shows that replacing individuals with essential societal roles, such as health workers, when they

are detected as being infectious, by susceptible individuals can accelerate disease transmission.

Note that we assume that the network structure of the population is static in the absence of disease. Depending on the disease of interest it would be interesting to consider a network that is dynamic also in the absence of infection, as would be appropriate for STI such as HIV to incorporate partner separation and formation over time. Superimposed on the dynamic network are then the dynamics that follow from social distancing (or other preventive measures). These are just a few important ways to modify and extend the social distancing model that we consider. As we find counterintuitive results already in the current model with relatively simple social distancing rules, it is difficult to understand how such extensions impact the epidemic, and certainly it would be interesting to investigate that in future work.

However, the aim of our paper is to show that rational individual-level preventive measures *can* have counter-intuitive consequences for the population-level. Public health interventions that aim at changing individual behaviour through social distancing could have adverse consequences, for example school closures could reduce social contacts between children in the school classes but may (partly) be replaced by social contacts outside of school. As our results show, it is not necessarily straightforward what effects such behaviour may have at the population level. These findings highlight the importance of modelling individual level behavioural changes in response to an epidemic to understand infectious disease dynamics.

Competing interests

We have no competing interests.

Authors' contributions

K.Y.L., F.B., D.S., and T.B. designed and performed research; K.Y.L. and T.B. analyzed data; F.B. and D.S. advised on analysis and interpretation; K.Y.L. and T.B. wrote the manuscript; and F.B. and D.S. contributed to writing the manuscript.

Funding

This work was partially supported by a grant from the Simons Foundation and was carried out as a result of the authors' visit to the Isaac Newton Institute for Mathematical Sciences during the programme Theoretical Foundations for Statistical Network Analysis in 2016 (EPSRC Grant Number EP/K032208/1). T.B. and K.Y.L. are supported by the Swedish Research Council (VR) Grant Number 2015-050153. This work was also supported by a grant from the Knut and Alice Wallenberg Foundation, which enabled F.B. to be a guest professor at the Department of Mathematics, Stockholm University.

Acknowledgements

We thank two anonymous referees for valuable comments that helped improved the manuscript. We thank A. Allard for bringing ref [23] to our attention.

References

1. Heesterbeek H, Anderson RM, Andreasen V, Bansal S, De Angelis D, Dye C, Eames KTD, Edmunds WJ, Frost SDW, Funk S, Hollingsworth TD, House T, Isham V, Klepac P, Lessler J, Lloyd-Smith JO, Metcalf CJE, Mollison D, Pellis L, Pulliam JRC, Roberts MG, Viboud C, Isaac Newton Institute IDD Collaboration 2015. Modeling infectious disease dynamics in the complex landscape of global health. *Science* **347**(6227):aaa4339.
2. Rubin GJ, Amlot R, Page L, Wessely S. 2009 Public perceptions, anxiety, and behavior change in relation to the swine flu outbreak: cross sectional telephone survey. *BMJ* **339**:b2651.
3. Jones JH, Salathe M. 2009 Early assessment of anxiety and behavioral response to novel swine origin influenza A (H1N1). *PLOS ONE* **4**:e8032.
4. Bayham J, Kuminoff NV, Gunn Q, Fenichel EP. 2015 Measured voluntary avoidance behaviour during the 2009 A/H1N1 epidemic. *Proc R Soc B* **282**:20150814.
5. Fast SM, Mekaru S, Brownstein JS, Postlethwaite TA, Markuzon N. 2015 The role of social mobilization in controlling Ebola Virus in Lofa County, Liberia. *Plos Curr* 2015, May 15.
6. Funk S, Ciglenecki I, Tiffany A, Gignoux E, Camacho A, Eggo RM, Kucharski AJ, Edmunds WJ, Bolongei J, Azuma P, Clement P, Alpha TS, Sterk E, Telfer B, Engel G, Parker LA, Suzuki M, Heijnenberg N, Reeder B. 2017 The impact of control strategies and behavioural changes on the elimination of Ebola from Lofa County, Liberia. *Phil Trans R Soc B* **372**:20160302.
7. Funk S, Salathé M, Jansen VAA. 2010 Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J R Soc Interface* **7**(50):1247–1256.
8. Manfredi P, d’Onofrio A, editors. *Modeling the interplay between human behavior and the spread of infectious diseases*. 2013 Springer-Verlag, New York.
9. Funk S, Bansal S, Bauch CT, Eames KTD, Edmunds WJ, Galvani AP, Klepac P. 2015 Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. *Epidemics* **10**:21–25.
10. Verelst F, Willem L, Beutels P. 2016 Behavioural change models for infectious disease transmission: a systematic review (2010-2015). *J R Soc Interface* **13**:20160820.
11. Newman MEJ. 2010 *Networks: an introduction*. Oxford University Press, Oxford.
12. Danon L, Ford AP, House T, Jewell CP, Keeling MJ, Roberts GO, Ross JV, Vernon MC. 2011 Networks and the epidemiology of infectious disease. *Interdiscip*

Perspect Infect Dis **2011**:1–28.

13. Bollobás B. 2001 *Random Graphs*. Academic Press, New York.
14. Molloy M, Reed B. A critical point for random graphs with a given degree sequence. 1995 *Random Structures and Algorithms* **6**:161–179.
15. Newman MEJ, Strogatz SH, Watts DJ. 2001 Random graphs with arbitrary degree distributions and their applications. *Phys Rev E* **64**:026118.
16. Durrett R. *Random graph dynamics*. 2006 Cambridge University Press, Cambridge.
17. Janson S. 2009 The probability that a random multigraph is simple. *Comb Probab Comput* **18**:205–225.
18. Ball F, Sirl D, Trapman P. 2009 Threshold behaviour and final outcome of an epidemic on a random network with household structure. *Adv Appl Probab* **41**:765–796.
19. Leskovec J, Krevl A. SNAP Datasets: Stanford large network dataset collection. Accessed: <http://snap.stanford.edu/data>, June 2014.
20. Britton T, Juher D, Saldana. 2016 A network epidemic model with preventive rewiring: comparative analysis of the initial phase. *Bull Math Biol* **78**:2427–2454.
21. Meloni S, Perra N, Arenas A, Gómez S, Moreno Y, Vespignani A. 2011 Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific reports* **1**:62.
22. Nicolaides C, Cueto-Felgueroso L, Juanes R. 2013 The price of anarchy in mobility-driven contagion dynamics. *J R Soc Interface* **10**:20130495.
23. Scarpino SV, Allard A, Hébert-Dufresne L. 2016 The effect of a prudent adaptive behaviour on disease transmission. *Nature Physics* **12**:1042-1047

Supplementary Information

Individual preventive social distancing during an epidemic may have negative population-level outcomes

Ka Yin Leung^{a,1}, Frank Ball², David Sirl², and Tom Britton¹

5th June 2018

This supplementary information accompanies the manuscript ‘Individual preventive social distancing during an epidemic may have negative population-level outcomes’. It contains details on mathematical results, additional figures, and summary statistics for the two real-world networks.

In Section S1 analytical results and additional figures are presented for the social distancing model on the configuration network model. In Section S1.3 we provide a proof that the final size can increase with increasing social distancing rate ω by considering an asymptotic lower bound for the final size. In ?? we provide some additional simulation studies on different configuration networks. In Section S2 we derive the epidemic threshold parameter R_* for the social distancing model on the clique network, the so-called clique reproduction number. We show that social distancing can move R_* from below to above the threshold value of one, thereby leading to a positive probability of a major outbreak to occur. In Section S3 properties of the two real-world networks are listed. In Section S4 we show that negative population level effects can occur for a range of α values large enough, where ‘large enough’ depends on the network structure under consideration and other parameter value choices. Finally in Section S5 we describe the technicalities of the simulation studies.

Section S1 Configuration network

Section S1.1 Notation and construction

The (Newman-Strogatz-Watts version of the) configuration model is defined in the main text in Section 2.2.1. The degree D of an individual in the network is distributed as the pre-defined distribution $\{p_d\}$. The mean and variance of D are denoted by μ_D and σ_D^2 , respectively. In the configuration network construction a given half-edge is d times as likely to be paired with a given individual with degree d than with a given individual with degree 1. Therefore, a neighbour has size-biased degree \tilde{D} , where $P(\tilde{D} = d) = dp_d/\mu_D$, $d = 1, 2, \dots$. The configuration model is well studied, e.g. [7]. It is for instance known that, as the population size $n \rightarrow \infty$, there will be zero connected components that have size of exact order n if $\mu_D + \sigma_D^2/\mu_D - 1 \leq 1$, and precisely

^akayin.leung@math.su.se

¹Department of Mathematics, Stockholm University, 106 91 Stockholm, Sweden

²School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK

one connected component of exact order n if $\mu_D + \sigma_D^2/\mu_D - 1 > 1$. Since there is at most one component that has size of exact order n this component is often called the giant component. Further, if the degrees are not small, then the giant often makes up close to all individuals. If for example all individuals have the same degree $d > 2$, then all nodes are part of the giant component with a probability tending to 1 as $n \rightarrow \infty$.

Section S1.2 Epidemic threshold parameter R_0

The basic reproduction number R_0 for the configuration network model with rewiring and dropping has been derived previously in [8]:

$$R_0 = \frac{\beta}{\beta + \omega + \gamma} \left(\mu_D + \frac{\sigma_D^2}{\mu_D} - 1 \right). \quad (\text{S.1})$$

From (S.1) we see that R_0 is a monotonically decreasing function of ω , i.e. any social distancing always reduces R_0 . In particular, if R_0 is larger than one in the baseline setting without any social distancing ($\omega = 0$), then social distancing reduces R_0 to below its epidemic threshold of one if and only if $\omega \geq \beta(\mu_D + \sigma_D^2/\mu_D - 1) - \gamma$. Therefore, social distancing is always beneficial at the beginning of an epidemic in the configuration network.

Section S1.3 Asymptotic lower bound for the final size

Assume that the degree distribution of the population only takes on values 0 and d , so $p_0 = 1 - p_d$ with $d > 2$. For ease of argument we consider the asymptotic lower bound first for the simpler setting of our model where $\gamma = 0$ and $\alpha = 1$, i.e. an SI infection (without recovery) and social distancing is always through rewiring to new individuals. We indicate later how the argument may be extended to the case when $\gamma = 0$ and $\alpha \in (0, 1]$. By a continuity argument it follows that, for any $\alpha \in (0, 1]$, the conclusions also hold for all sufficiently small $\gamma > 0$. The degree of an individual may change over the course of an epidemic owing to rewiring. We say that an individual has ‘original degree’ 0 or d if that was the degree of the individual in the configuration-network construction before the epidemic.

We consider R_0 as a function of ω . In the setting of this section, $R_0 = R_0(\omega) = \beta(d-1)/(\beta+\omega)$ (in the early stages of the epidemic, all newly infected individuals have degree d). Note that $R_0(0) = d - 1$, i.e. in an SI epidemic without rewiring, a newly infected individual transmits infection to all its susceptible neighbours. Note that $R_0(0) > 1$ since we have assumed that $d > 2$. Let $\bar{Z}_n(\omega)$ denote the final fraction of the population that gets infected in the model with population size n and rewiring rate ω . If $\omega = 0$, then only transmission can occur (recall that we consider $\gamma = 0$), so an index case will generate an outbreak of the size equal to the size of the connected component this index case is part of. Asymptotically, as population size $n \rightarrow \infty$, as mentioned earlier, all individuals with degree d will be part of the giant component of the network so the giant component consists of a fraction p_d of the entire population. Therefore, as n tends to infinity, $\bar{Z}_n(0)$ tends to a two-point distribution $\bar{Z}(0)$ with $P(\bar{Z}(0) = 0) = p_0 = 1 - P(\bar{Z}(0) = p_d)$. That is, in the setting without recovery, the asymptotic final fraction infected is either zero (if the index case has degree 0) or p_d (if the index case has degree d).

Note that for large enough ω , $R_0(\omega) < 1$. We consider $\omega > 0$ sufficiently small such that $R_0(\omega) > 1$. Furthermore, we assume that a major outbreak occurs. Therefore we assume that the index case has degree d (if the index case has degree 0, then the final fraction $\bar{Z}_n(\omega)$ tends to zero as n tends to infinity). Then the final fraction $\bar{Z}_n(\omega)$ tends in probability to $\tau(\omega)$ as $n \rightarrow \infty$. We show that the relative final size $\tau(\omega)$ of a major outbreak can increase for ω compared to the baseline $\omega = 0$ model. We do so by giving a lower bound $\tilde{\tau}(\omega)$ for $\tau(\omega)$. This lower bound

is obtained by considering a model with slightly different social distancing rules. Suppose that a susceptible individual distances him/herself from its infectious neighbour. Then he/she tries to rewire to a randomly chosen individual v in the population. If v is an individual with original degree d , then no connection is made and the existing connection is dropped instead. On the other hand, if v has original degree 0, then (i) a connection is made as in the original model and (ii) v is not allowed to transmit infection to an individual with original degree d . Note that all these modifications make infection less likely, so the final size of this modified model is less than that of the original social distancing model. Furthermore, note that in the baseline model without rewiring ($\omega = 0$) the final size of the original model is equal to the lower bound obtained from the modified model.

Mean final size

Previous work [5, Theorem 6 and Proposition 4] yields that the deterministic final fraction of the population with original degree d that becomes infected in the epidemic is $\rho = 1 - z^d$ where $z \in (0, 1)$ is the unique solution to

$$(\beta + \omega)z - \omega - \beta z^{d-1} = 0 \quad (\text{S.2})$$

(note that by considering only the original degree d individuals, the probability generating function in [5] is given by $f_D(s) = s^d$ and $\mu_D = d$). Let $\theta = \omega/(\beta + \omega)$ denote the probability that social distancing occurs before transmission for a connection between a susceptible and an infectious individual. Any original degree d individual that is exposed to at least one infectious individual will rewire at least once with probability θ , so the fraction of the original degree d population that rewires at least once is at least $\rho\theta$. These will include individuals that are not subsequently infected. However, since the fraction of the original degree d population that avoid infection is $1 - \rho$, the fraction of that population which rewire at least once and are subsequently infected is at least $\rho\theta - (1 - \rho)$. The probability that such a rewiring is to an original degree 0 individual is p_0 and, given it is to an original degree 0 individual, the probability that individual itself rewires before he/she is infected is $1 - \theta$. Thus the fraction of the original degree d population that rewires and transmits infection to an original degree 0 individual is at least

$$\phi = (\rho\theta - (1 - \rho))p_0(1 - \theta). \quad (\text{S.3})$$

Asymptotically, the probability that an individual of degree 0 escapes infection is at most $(1 - 1/(np_0))^{\phi np_d} \rightarrow e^{-\phi p_d/p_0}$ as $n \rightarrow \infty$, since approximately at least $n\phi p_d$ individuals rewire and transmit infection to an original degree 0 individual, and for each such rewiring event the individual rewired to is chosen uniformly at random from the np_0 original degree 0 individuals. Therefore, a lower bound for the fraction of the population that originally had degree 0 and get infected is $p_0(1 - e^{-\phi p_d/p_0})$. Hence, the total fraction of the population that eventually get infected is, in the limit as $n \rightarrow \infty$, at least

$$\tilde{\tau}(\omega) = p_d \rho + p_0(1 - e^{-\phi p_d/p_0}). \quad (\text{S.4})$$

The latter yields an asymptotic lower bound for the final fraction infected in the SI epidemic on the configuration network with rewiring rate ω (i.e. where $\gamma = 0$, $\alpha = 1$). We consider the first order approximation of (S.4) in ω to show that the lower bound for the relative final size is exact when $\omega = 0$ and increasing for small ω , thus the final size of the original model also increases for small ω . We approximate ρ by considering (S.2). Denote the left-hand side of (S.2) by $F(z(\omega), \omega)$, so $F(z(\omega), \omega) = 0$. Then $\partial F/\partial z \cdot z'(\omega) + \partial F/\partial \omega = 0$, which yields $z'(\omega) = -(z - 1)/(\beta + \omega - \beta(d - 1)z^{d-2})$. Hence $z(\omega) = z(0) + z'(0)\omega + o(\omega) = \omega/\beta + o(\omega)$,

and we can approximate $\rho = 1 - (\omega/\beta)^d + o(\omega^d)$ for $\omega > 0$ small enough. Next, recalling that $\theta = \omega/(\beta + \omega)$ we make the approximation

$$\begin{aligned}\phi &= [(1 - (\omega/\beta)^d)\theta - (\omega/\beta)^d + o(\omega^d)] p_0(1 - \theta) \\ &= p_0\omega/\beta + o(\omega)\end{aligned}$$

as $\omega \downarrow 0$. The first order approximation of (S.4) is therefore

$$\tilde{\tau}(\omega) = p_d(1 + p_0\omega/\beta) + o(\omega).$$

Recall that the limiting mean fraction infected in the event of a large outbreak in the model with no rewiring ($\omega = 0$) is p_d , i.e. this approximation is exact when $\omega = 0$. Thus, since $p_d(1 + p_0\omega/\beta) \geq p_d$ for all $p_0, \omega > 0$, we find that the final fraction infected in case of a major outbreak is increased for all sufficiently small but strictly positive rewiring rates $\omega > 0$ for the modified model. In other words, the asymptotic lower bound for the relative final size given a major outbreak is increasing in ω for small $\omega > 0$.

Mean final size of a major outbreak

For the expected relative final size we have

$$E(\bar{Z}(\omega)) = \tau(\omega)P_\omega(\text{major outbreak}) \geq p_d(1 + \frac{p_0}{\beta}\omega)P_\omega(\text{major outbreak}) + o(\omega).$$

since the expected relative final size of a minor epidemic is 0. In order to show that also $E(\bar{Z}(\omega))$ has a lower bound that is increasing in ω for sufficiently small ω , it suffices to show that $P_\omega(\text{major outbreak})$ does not decrease too fast. We show that $1 - P_\omega(\text{major outbreak}) = P_\omega(\text{minor outbreak}) \leq p_0 + p_d\omega^{d-1} + o(\omega^{d-1})$ or equivalently that

$$P_\theta(\text{minor outbreak}) \leq p_0 + p_d\theta^{d-1} + o(\theta^{d-1}). \quad (\text{S.5})$$

Note that there is a minor outbreak if either the index case has degree 0 (with probability p_0), or the index case has degree d (with probability $p_d = 1 - p_0$) but the outbreak that takes place amongst the individuals with degree d is a minor outbreak. Consider the configuration model without the degree 0 individuals. We modify the model by letting the index case have degree $d-1$ and all other individuals have degree d . Consequently, *all* individuals, including the index case, have $d-1$ susceptible neighbours when newly infected in the beginning of the epidemic. This modified model has a probability of a minor outbreak that is larger than the original model (and is exact in the case $\theta = 0$), i.e. $P_\theta(\text{minor outbreak}) \leq P_\theta(\text{minor outbreak of the modified model})$. Let $\pi(\theta)$ denote the probability of a minor outbreak for the modified model, given that the index case is of degree $d-1$ (with probability $p_d = 1 - p_0$). By conditioning on the number of neighbours that avoid infection in the first generation, i.e. the number of neighbours that rewire away before infection, we find the consistency equation

$$\pi(\theta) = \sum_{j=0}^{d-1} \pi(\theta)^{d-1-j} \binom{d-1}{j} \theta^j (1-\theta)^{d-1-j} = (\theta + \pi(\theta)(1-\theta))^{d-1}.$$

Note that $\pi(0) = 0$ since a major outbreak will occur without any rewiring. By the Taylor expansion of $\pi(\theta)$ about $\theta = 0$, we find that $\pi(\theta) = \theta^{d-1} + o(\theta^{d-1})$, proving (S.5). Since $E(\bar{Z}(0)) = p_d$, we have a lower bound for the expected relative final size that is increasing for sufficiently small $\omega > 0$, and is exact for the baseline model $\omega > 0$ without rewiring.

Allowing dropping

Suppose that $\alpha \in (0, 1)$. Then since in the model with modified social distancing all attempted rewirings to an original degree d individual results in the edge being dropped, the only change in the argument leading to (S.4) is that the expression for ϕ in (S.3) becomes

$$\phi = (\rho\theta - (1 - \rho))p_0(1 - \theta)\alpha.$$

It follows that $\phi = p_0\alpha\omega/\beta + o(\omega)$ as $\omega \downarrow 0$ and, since $\alpha > 0$, the asymptotic lower bound for the relative final size given a major outbreak is again increasing in ω for small $\omega > 0$. Finally, note that the probability of a major outbreak is independent of α , so the above argument shows that the expected relative final size is also increasing in ω for small $\omega > 0$. These results are illustrated with simulation studies in Fig. S4

Section S1.4 The number of index cases

We consider the social distancing model on the configuration network with degree distribution $\{p_d\}_{d=0}^{10}$ where $p_d = c/(d+1)$, $d = 0, 1, \dots, 10$, with $c = 0.331$ the normalization constant. In the main text, each epidemic is initiated with ten initially infected individuals that are randomly chosen from the population. The rationale behind this choice is to ensure sufficiently many epidemics result in major outbreaks. Owing to the degree distribution of the population, a large fraction of the population has degree 0 or degree 1. An individual with degree 0 has no connections and is therefore unable to transmit infection in the population. Also an individual with degree 1 is unlikely to generate a major outbreak. In Fig. S1 we compare the final epidemic size for the case that each epidemic starts with 10 randomly chosen infectives to the case that each epidemic starts with only 1 randomly chosen infective.

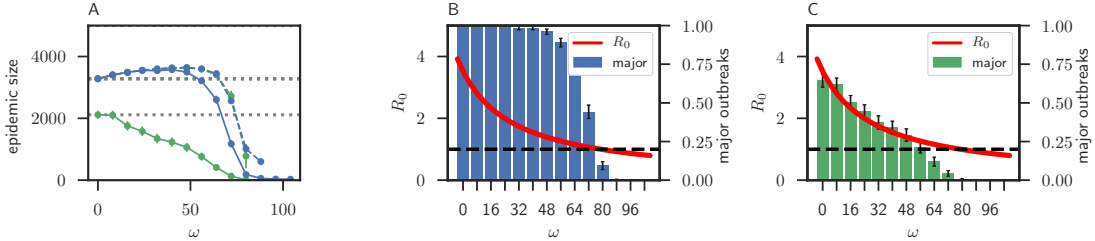


Figure S1: The effect of social distancing on the final size of epidemics with different numbers of initially infected individuals. (A) Average final size compared to the average final size in the baseline model (with 95% CI) over all outbreaks (solid line) and restricted to major outbreaks (dashed line); dotted horizontal lines are at the size of the epidemic when $\omega = 0$, for reference. (B) R_0 as a function of ω (red line) and fraction of outbreaks resulting in a major outbreak for the setting of 10 initially infectives (with 95% CI). (C) R_0 as a function of ω (red line) and fraction of outbreaks resulting in a major outbreak for the setting of 1 initially infective (with 95% CI). Black dashed line at $R_0 = 1$ indicates the threshold value for the beginning of the epidemic. We consider the configuration network model with degree distribution $\{p_d\}_{d=0}^{10}$ where $p_d = c/(d+1)$, $d = 0, 1, \dots, 10$, with $c = 0.331$ the normalization constant. The population size is 5000 and 500 simulation runs are performed for each value of ω . Parameter values are $\beta = 20, \gamma = 0.2, \alpha = 0.9$. The index cases are randomly chosen in the entire population. The blue curves show the results for 10 initially infected individuals (results main text Fig. 1) whereas the green curve show the results for one initially infected individual.

Note that the effects of the number of initial infectives as shown in Fig. S1 are not surprising.

The average final size over all outbreaks is much smaller in the setting where epidemics are initiated by one initial infective individual. Since this individual is randomly chosen over the entire population, it is quite likely that he/she has degree 0 or 1. In such cases, epidemics will (most likely) die out (Fig. S1C). By considering 10 initial infectives instead, it becomes more likely that an epidemic outbreak occurs (Fig. S1B). This is both because there are simply more infectives to start with, making it less likely that the epidemic dies out owing to stochastic effects, and it being more likely that some initial infectives have higher degree, enabling further transmission in the population. On the other hand, the average final size in case of a major outbreak does not depend on the initial number of infectives (Fig. S1A, blue and green dashed lines).

Section S2 The clique-network model

In the main text as well as in this supplement we consider the clique-network model where all cliques consist of exactly three individuals. The reason for this choice is that it allows us to illustrate our point that social distancing can have a negative effect on the initial stages of an outbreak by increasing the epidemic threshold parameter R_* . At the same time, the choice of cliques of size three simplifies the bookkeeping in the calculations below. Moreover, it yields an explicit expression for the clique reproduction number R_* . However, in principle similar calculations can be done also for clique sizes larger than three, or having variable clique sizes.

The clique network with variable clique sizes can be constructed as follows (see also [2, 3] where the same model is referred to as household-network models). A clique-size distribution $\{\pi_h\}$ is predefined that describes the sizes of cliques in the population. The population can be divided into cliques with this distribution by drawing random sizes h_1, h_2, \dots independently from $\{\pi_h\}$, then labelling all individuals 1 up to n , and letting the first h_1 individuals make up clique 1, the next h_2 individuals make up clique 2, and so on. In the final network, individual 1 is then connected to all individuals he/she is connected to from the construction of the configuration model and individuals 2 up to h_1 from the clique construction, and so on.

We denote the degree distribution for the global network connections by $\{p_d\}$, and the mean and variance by μ_D and σ_D^2 , respectively. We also use the size-biased degree distribution $\{\tilde{p}_d\}$ with $\tilde{p}_d = dp_d/\mu_D$. We let $\tilde{\mu}_D$ denote the mean of the size-biased degree distribution, where $\tilde{\mu}_D = \sum_{d=1}^{\infty} d\tilde{p}_d = \sum_{d=1}^{\infty} d^2 p_d/\mu_D = \mu_D + \sigma_D^2/\mu_D$.

Section S2.1 Derivation of the epidemic threshold parameter R_*

We consider the clique reproduction number R_* that can be interpreted as the expected number of secondary cliques generated by one typical newly infected clique at the beginning of an epidemic [6, 1, 4]. So, rather than considering individuals as the units of interest we consider cliques. The clique reproduction number R_* satisfies the desired threshold behaviour that, as the number of cliques tends to infinity, there is a strictly positive probability of a major outbreak if $R_* > 1$ and only minor outbreaks occur if $R_* \leq 1$.

At the start of an epidemic there are only few infected individuals and most individuals are susceptible. Therefore, a newly infected clique at the beginning of an epidemic has most likely only one newly infected individual that got infected by a neighbour outside his/her own clique. In the SIR epidemic on the clique-network with social distancing, there are exactly two types of newly infected clique: the first infected individual in the clique was infected through (1) a global network neighbour or (2) a rewired network neighbour. Note that in case (1) the individual has $d - 1$ susceptible global connections with probability \tilde{p}_d while in case (2) the individual has d susceptible global connections with probability p_d (note that the probability that an individual

is rewired to more than once is negligible in the early stages). The other two clique members have d susceptible global connections with probability p_d .

The clique reproduction number R_* is given by the dominant eigenvalue of the 2×2 matrix $K = (K_{ij})_{i,j=1,2}$, where K_{ij} is the expected number of cliques of type j generated by one newly infected clique of type i . In the remainder of this section we derive expressions for the K_{ij} by considering all possible events that can occur within the clique. We denote the clique member that was infected from outside of the clique by u_* . The other two clique members are denoted by u_1 and u_2 . We let

$$\theta_1 = \frac{\beta}{\beta + \omega + \gamma}, \quad \theta_2 = \frac{\omega}{\beta + \omega + \gamma}, \quad \theta_3 = 1 - \theta_1 - \theta_2 = \frac{\gamma}{\beta + \omega + \gamma},$$

so that θ_1 is the probability that transmission occurs from an infectious individual to a given susceptible connection before recovery or social distancing, θ_2 is the probability that social distancing occurs before transmission or recovery, and θ_3 is the remaining probability (of recovery before either event).

The K_{ij} can be derived directly from the above interpretation. First of all, note that $K_{12} = K_{22}$, i.e. the expected number of secondary cliques of type 2 generated by a newly infected clique does not depend on the type of that clique as the global network degree of u_* does not play a role. A clique member within a clique can transmit infection to individuals in another clique through a rewired edge if (i) there is at least one clique edge that is rewired away to a new individual in the population, (ii) the individual u_1 that rewires away becomes infected by a clique member, and (iii) u_1 transmits infection along its rewired clique connection. Since we need both that clique connections rewire to new cliques and the individuals that rewire away to become infected, the only possibility for secondary cliques of type 2 to be generated is if there is exactly one rewired clique connection in the index clique.

When u_* is newly infected and the other two clique members u_1 and u_2 are (still) susceptible, the following events can occur: (i) u_* recovers (at rate γ), (ii) u_* transmits to u_1 or u_2 (at rate 2β), or (iii) u_1 or u_2 distances itself from u_* (at rate 2ω). So we find that the probability that u_* transmits to u_1 is

$$\pi = \frac{\beta}{\gamma + 2\beta + 2\omega}. \tag{S.6}$$

The other probabilities in the derivations below can be derived in similar manner and we leave out the details. We find the following two possibilities.

- u_* transmits to u_1 with probability π , then u_2 rewires one of its clique connections from either u_* or u_1 (with probability $2\alpha\omega/(2\beta + 2\omega + 2\gamma)$), after which u_2 becomes infected through its clique connection that was not rewired away (with probability θ_1). Finally, u_2 transmits to a global neighbour through its rewired clique edge (with probability θ_1)
- u_1 rewires away from u_* to a new individual (with probability $2\alpha\omega/(2\beta + \gamma + 2\omega)$), then u_* transmits to u_2 (with probability θ_1). Next, u_2 transmits infection to u_1 (with probability θ_1), and u_1 transmits to a global neighbour through its rewired clique edge (with probability θ_1)

Therefore,

$$K_{12} = \frac{4\alpha\omega}{(2\beta + \gamma + 2\omega)}\theta_1^3 = K_{22}. \tag{S.7}$$

K_{11} and K_{21} are a bit more involved as there are more possibilities to take into account. Note that the only difference between the two K_{i1} is in the degree of u_* . We note that the

expected number of newly infected cliques generated by network edges of u_* is the probability θ_1 of transmission through a network connection (before recovery of u_* or rewiring of network edges) times the expected number of susceptible network neighbours of u_* . If u_* is of type 1, then the expected number of susceptible network neighbours is $\tilde{\mu}_D - 1$. If u_* is of type 2, then this expected number is μ_D .

Next, we note that clique members u_1 and u_2 are interchangeable in the sense that they both have network degree d with probability p_d , and all three clique members u_1 , u_2 , and u_* are connected to each other. If individual u_1 gets infected, then the expected number of susceptible network neighbours it has is μ_D . The expected number of secondary cases an infectious clique member u_1 generates is then $\theta_1\mu_D$. We derive the probability that u_1 becomes infected by taking into account all relevant events and the order in which they occur.

The following events can occur (with corresponding probabilities):

- u_* transmits to u_1 with probability π
- u_* transmits to u_2 with probability π , after which u_* or u_2 transmits to u_1 with probability $\frac{2\beta}{2\gamma+2\beta+2\omega} = \theta_1$.
- u_* transmits to u_2 with probability π , then u_1 distances itself from either u_* or u_2 with probability θ_2 , and finally u_1 becomes infected by the clique member it did not distance itself from with probability θ_1
- u_* transmits to u_2 with probability π , then u_* or u_2 recovers with probability θ_3 , and finally u_1 becomes infected by the clique member that is still infectious with probability θ_1
- u_2 distances itself from u_* with probability $\frac{\omega}{\gamma+2\beta+2\omega}$ after which u_* transmits to u_1 with probability θ_1
- u_1 distances itself from u_* with probability $\frac{\omega}{\gamma+2\beta+2\omega}$, then u_* transmits to u_2 with probability θ_1 , and finally u_2 transmits to u_1 with probability θ_1

The probability that u_2 becomes infected is obtained by interchanging the names u_1 and u_2 , so we simply multiply by a factor 2 in the expected number of infected cliques generated by u_1 and u_2 .

Putting these pieces together, we find that the expressions for the K_{ij} , $i, j = 1, 2$, are as follows:

$$\begin{aligned}
K_{11} &= \theta_1(\tilde{\mu}_D - 1) + 2\theta_1\mu_D \left(\frac{\beta}{\gamma + 2\beta + 2\omega} + \frac{\beta}{\gamma + 2\beta + 2\omega}\theta_1 + \frac{\beta}{\gamma + 2\beta + 2\omega}\theta_1\theta_2 \right. \\
&\quad \left. + \frac{\beta}{\gamma + 2\beta + 2\omega}\theta_3\theta_1 + \frac{\omega}{\gamma + 2\beta + 2\omega}\theta_1 + \frac{\omega}{\gamma + 2\beta + 2\omega}\theta_1^2 \right) \\
&= \theta_1(\tilde{\mu}_D - 1) + 2\theta_1\mu_D \left(\frac{\beta}{\gamma + 2\beta + 2\omega}(1 + \theta_1 + \theta_1\theta_2 + \theta_1\theta_3) + \frac{\omega}{\gamma + 2\beta + 2\omega}\theta_1(1 + \theta_1) \right) \\
K_{21} &= \theta_1\mu_D + 2\theta_1\mu_D \left(\frac{\beta}{\gamma + 2\beta + 2\omega}(1 + \theta_1 + \theta_1\theta_2 + \theta_1\theta_3) + \frac{\omega}{\gamma + 2\beta + 2\omega}\theta_1(1 + \theta_1) \right)
\end{aligned}$$

Note that the sum $2\frac{\omega}{(\gamma+2\beta+2\omega)}\theta_1^2 = \frac{\beta}{\gamma+2\beta+2\omega}\theta_1\theta_2 + \frac{\omega}{\gamma+2\beta+2\omega}\theta_1^2$ of the two terms $\frac{\beta}{\gamma+2\beta+2\omega}\theta_1\theta_2$ and $\frac{\omega}{\gamma+2\beta+2\omega}\theta_1^2$ in K_{11} and K_{21} is the probability of social distancing of an initially susceptible clique member that later becomes infected.

Finally, the clique reproduction number R_* is the dominant eigenvalue of the 2×2 matrix $K = (K_{ij})_{i,j=1,2}$. Therefore, we have an explicit expression of R_* in terms of the model parameters:

$$R_* = \frac{\text{tr} + \sqrt{\text{tr}^2 - 4\det}}{2}, \quad (\text{S.8})$$

where $\text{tr} = K_{11} + K_{22}$ and $\det = K_{11}K_{22} - K_{12}K_{21}$ denote the trace and determinant of the matrix K , respectively.

Section S2.2 Social distancing can increase R_* but this depends on the network structure

We consider two degree distributions for the degree D of the global network, and using (S.8) for R_* , show that R_* can increase for sufficiently small distancing rate $\omega > 0$ and recovery rate $\gamma > 0$. We consider a third choice for D to show that this need not always be the case. Assume that $\gamma = 0$ and $\alpha = 1$ and write $R_* = R_*(\omega)$.

First, suppose that $p_1 = p = 1 - p_0$. Then $\mu_D = p$ and $\mu_{\tilde{D}-1} = 0$. Then $R_*(0) = 2p$ and $R'_*(0) = (3 - 2p)/\beta$. In particular, if $p = 1/2$, then $R_*(0) = 1$, and $R'_*(0) > 0$. Therefore, for sufficiently small $\gamma > 0$ and sufficiently large $\alpha < 1$, an increasing rewiring rate $\omega > 0$ can push the epidemic threshold parameter R_* from below to above the threshold value of one as shown numerically in Fig. 2A and Fig. 2C of the main text.

Next, suppose that D is Poisson with mean μ_D . Then $\tilde{D} - 1$ is also Poisson with mean μ_D . Then $R_*(0) = 3\mu_D$, and $R'_*(0) = (2 - 3\mu_D)/\beta$. In particular, if $\mu_D = 1/3$, then $R_*(0) = 1$, and $R'_*(0) > 0$. Thus, for sufficiently small $\gamma > 0$ and sufficiently large $\alpha < 1$ and μ_D close to $1/3$, an increasing rewiring rate ω can push the epidemic threshold parameter R_* from below to above the threshold value of one as shown numerically in Fig. S2A.

Note that the derivations above do not rely on the specific choice $\alpha = 1$. Indeed, we have an analytical expression (S.8) for R_* that is explicit in terms of the model parameters so we can very well do the same derivations for some fixed value of α . By doing so, we can also derive a threshold for α . For the model with $p_1 = p = 1 - p_0$, one gets $\text{sign}[R'_*(0)] = \text{sign}[3\alpha - 2p]$ so $R'_*(0) > 0$ if $\alpha > 2p/3$. In particular, if $p = 1/2$ then $R'_*(0) > 0$ if $\alpha > 1/3$. For the Poisson model, one gets $\text{sign}[R'_*(0)] = \text{sign}[2\alpha - 3\mu_D]$, so $R'_*(0) > 0$ if $\alpha > 3\mu_D/2$. In particular, if $\mu_D = 1/3$ then $R'_*(0) > 0$ if $\alpha > 1/2$. Note that this implies that the same conclusions hold for R_* for a much larger range of probabilities α than illustrated in Fig. 2 of the main manuscript. Concretely, for the model with $p_1 = p = 1 - p_0$, R_* can cross the threshold value of one for the range $1/3 < \alpha \leq 1$ and for the model with Poisson degree the range is $1/2 < \alpha \leq 1$. We illustrate these results in Fig. S5 for the model with degree distribution $p_1 = p = 1 - p_0$.

Finally, to show that the result for R_* does not hold in general, we consider the following distribution. Suppose that $p_d = 1$, for some fixed $d > 1$, so all individuals have degree d . Then $\mu_D = d$ and $\mu_{\tilde{D}-1} = d - 1$, whence $R_*(0) = 3\mu_D - 1$ and $R'_*(0) = -(1 - 12\mu_D + 9\mu_D^2)/(\beta(3\mu_D - 1))$. Note that $1 - 12\mu_D + 9\mu_D^2 = (3\mu_D - 2)^2 > 0$ as $d > 1$. We find that $R'_*(0) < 0$ for $d > 1$. This shows that rewiring decreases the epidemic threshold parameter R_* for sufficiently small $\omega > 0$ (and $\alpha < 1$, $\gamma > 0$) as illustrated in Fig. S2B.

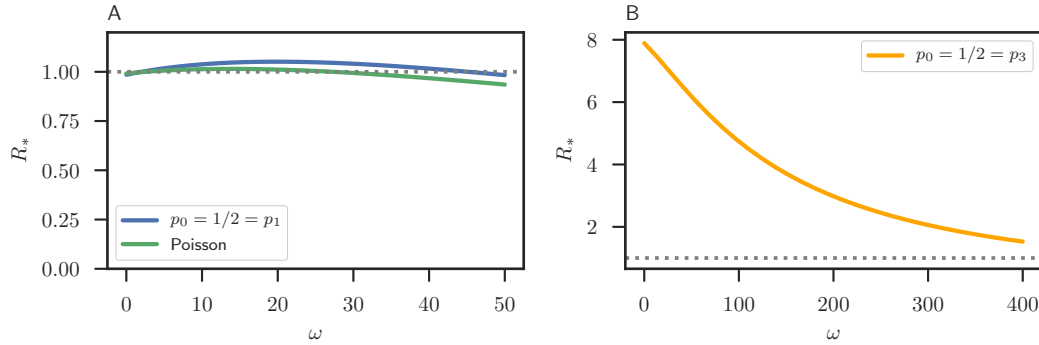


Figure S2: R_* for the clique-network model with cliques of size 3 with mean infectious period $1/\gamma = 5$ days, $\beta = 200$, $\alpha = 0.9$ (A) Social distancing can increase the epidemic threshold parameter R_* from below to above its threshold value of one for clique-network models. Illustration with two-point degree distribution with $p_0 = 1/2 = p_1$ (blue line) and Poisson degree distribution with mean $\mu_D = 0.335$ (green line) (B) Social distancing does not have to increase the epidemic threshold parameter R_* . Illustration with degree distribution $p_3 = 1$.

Section S3 Real-world networks

Section S3.1 Properties of the real-world networks

We summarize the most important characteristics of the real-world networks for our purpose, other network summary statistics and more details are found in [9].

The ‘arXiv General Relativity’ collaboration network describes scientific collaborations between authors that submitted papers to the arXiv in the General Relativity and Quantum Cosmology category. Edges between nodes represent two co-authors that have written a paper together. There are 5241 nodes and 14484 edges in the network. In total there are 354 connected components that make up the network and the largest connected component covers a fraction of 0.793 nodes and 0.926 edges. The minimum and maximum degree in the population are 1 and 81, with a mean of 5.53 and a variance of 62.7. The median degree is 3. There are in total 777 nodes with degree 3, and out of these nodes, 676 are part of the largest connected component. We choose the index case at random from these 676 nodes.

In the ‘Facebook social circles’ network, nodes are survey participants of the social network website Facebook that were using a specific app. Edges between nodes represent the ‘circles’ or ‘friends lists’ of those participants. There are 4039 nodes and 88234 edges. The largest network component is precisely the network itself. The minimum and maximum degree in the population are 1 and 1045, with a mean of 43.70 and a variance of 2748.44. The median degree in the population is 25, and there are 55 individuals with this median degree of 25. We choose an index case at random from this set of individuals.

Section S4 Probability α that social distancing is through rewiring rather than dropping

As previously noted, the setting that $\alpha = 0$, i.e. all social distancing is done through dropping of edges, is always beneficial for the population level. Therefore, only for α large enough, we

will find negative population level effects. However, whether α is large enough depends much on the network structure at hand and the choice of other parameter values. In this section we illustrate the analytical results of Section S1.3 and Section S2.2 that negative population level effects can occur for a range of α -values larger than the $\alpha = 0.9$ taken in the studies in the main manuscript. At the same time we show that these negative effects do not always occur.

First, in Fig. S3 we consider the scenario of Fig. 1 in the main manuscript, and compare the effect of $\alpha = 0.9$ to $\alpha = 0.5$. We see that when $\alpha = 0.5$, i.e. half of social distancing is through dropping of edges, then this has a positive population level effect in the setting presented in Fig. S3.

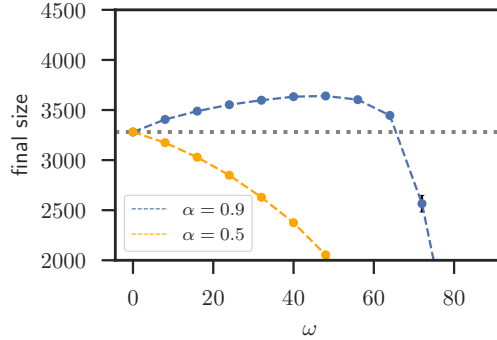


Figure S3: The effect of the probability α that social distancing is through rewiring rather than dropping on the final size of the epidemic given a major outbreak on the configuration network with degree distribution $\{p_d\}_{d=0}^{10}$ where $p_d = c/(d+1)$, $d = 0, 1, \dots, 10$, with $c = 0.331$ the normalization constant. Dotted horizontal line is for comparison with the size at $\omega = 0$. Model parameters are $1/\gamma = 5$ days, $\beta = 20$ days $^{-1}$, $\alpha = 0.9$ and 0.5 . Compare with Fig. 1 of the main manuscript (note that the y-axis is scaled differently for better comparison with α values). For each value of ω 500 epidemics are simulated. The epidemic is initiated with 10 index case that are chosen uniformly at random.

In Section Section S1.3 we proved that social distancing can have negative population level effects on the final size of the epidemic for $\alpha \in (0, 1]$. We illustrate this by considering an epidemic on a configuration network with degree distribution $p_0 = 1/2 = p_5$ for different values of α in Fig. S4. Note that generally the increase in the final size is smaller and occurs for a smaller range of social distancing rates $\omega > 0$ for smaller values of α .

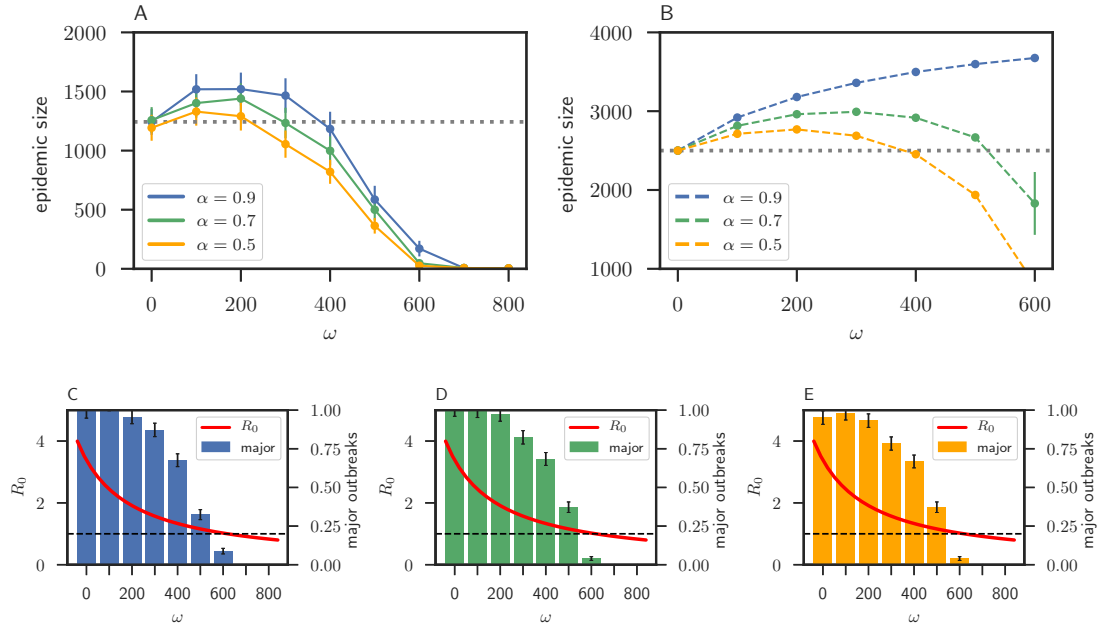


Figure S4: The effect of the probability α that social distancing is through rewiring rather than dropping on the final size of the epidemic over all epidemic outbreaks (A) and given a major outbreak (B) on the configuration network with degree distribution $p_0 = 1/2 = p_5$; dotted horizontal line is for comparison with the size at $\omega = 0$. Model parameters are $1/\gamma = 5$ days, $\beta = 200 \text{ days}^{-1}$, and $\alpha = 0.9$ (B), $\alpha = 0.7$ (C), $\alpha = 0.5$ (D) (note that the scaling of the y-axis for better comparison between different α values). For each value of ω 500 epidemics are simulated. The index case is chosen uniformly at random from degree 5 individuals.

Analysis in Section S2.2 shows that social distancing can push the epidemic threshold parameter R_* across the threshold value of one in a clique network for a range of α values. We illustrate this by considering the clique network with degree distribution $p_0 = 1/2 = p_1$ in Fig. S5. We see in Figs. S5A and B that increasing ω can lead to R_* crossing the threshold value of one for both $\alpha = 0.9$ and $\alpha = 0.5$. Note however that parameter values are such that for $\alpha = 0.5$, R_* remains very close to one, and therefore few epidemics result in major outbreaks. Furthermore, the negative population level effect is generally smaller for smaller values of α .

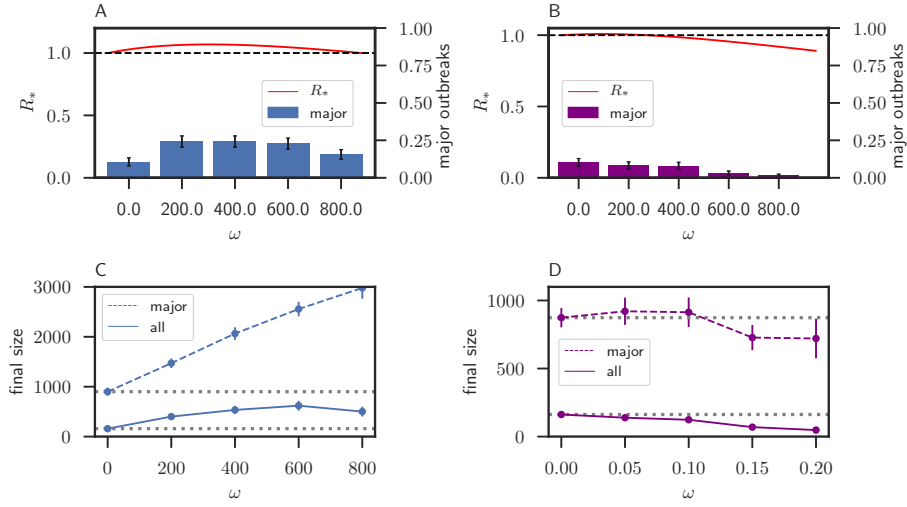


Figure S5: The effect of the probability α that social distancing is through rewiring rather than dropping on the final size of the epidemic given a major outbreak on the clique network with degree distribution $p_0 = 1/2 = p_1$; dotted horizontal lines are for comparison with the size at $\omega = 0$. Model parameters are $1/\gamma = 5$ days, $\beta = 4000$ days $^{-1}$, and (A, C) $\alpha = 0.9$, (B, D) $\alpha = 0.5$ (note that the y-axis is scaled differently in Figs. C and D). For each value of ω 500 epidemics are simulated. The index case is chosen uniformly at random.

In Fig. S6 we find that the final size given a major outbreak can increase for small enough social distancing rates ω under different rewiring probabilities α for an epidemic on the arXiv collaboration network. Note however that the range of social distancing rates ω for which this occurs, and the extent to which the final size can increase depend on the value of α . In contrast, in Fig. S7 we find that the probability α has little effect on the final size given a major outbreak for an epidemic on the facebook social circles network.

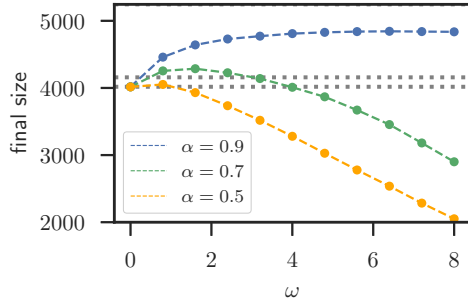


Figure S6: The effect of the probability α that social distancing is through rewiring rather than dropping on the final size of the epidemic given a major outbreak on the arXiv collaboration network; dotted horizontal lines are for the size of the giant component (top) and comparison with the size at $\omega = 0$. Model parameters are $1/\gamma = 5$ days, $\beta = 2 \text{ days}^{-1}$, and $\alpha = 0.9, 0.7$, and 0.5 . For each value of ω 500 epidemics are simulated. The index case is chosen uniformly at random from the sub-population of individuals that has median degree and are part of the largest connected component of the network. Compare with Fig. 3 of the main manuscript (note that the y-axis is scaled from 2000 to 5241 for better comparison between the α).

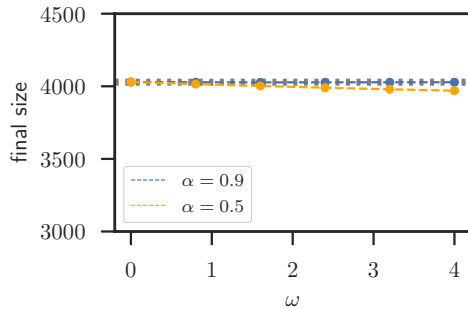


Figure S7: The effect of the probability α that social distancing is through rewiring rather than dropping on the final size of the epidemic given a major outbreak on the facebook social circles network; dotted horizontal lines are for the size of the network (top) and comparison with the size at $\omega = 0$. Model parameters are $1/\gamma = 5$ days, $\beta = 2 \text{ days}^{-1}$, and $\alpha = 0.9$, and 0.5 . For each value of ω 500 epidemics are simulated. The index case is chosen uniformly at random from the sub-population of individuals that has median degree. Compare with Fig. 4 of the main manuscript (note that the y-axis is scaled from 2000 to 4039 for better comparison between the α).

Section S5 Simulation studies of the social distancing model

We are interested the epidemic final size and the probability of a major outbreak for the social distancing model applied to (i) the configuration network model, (ii) the clique network model, and (iii) two real-world networks. The social distancing model is mathematically challenging to analyse. Although we have some analytical results (presented in Sections Section S1 and Section S2), e.g. for the threshold parameter R_0 for the configuration network model and R_* for the

clique network model, other quantities of interest, i.e. the final size and the probability of a major outbreak we have not managed to characterize mathematically. Moreover, on large real-world networks it is impractical to exactly compute the quantities we are interested in. The results in this text are obtained through simulation methods.

Given the structure of the networks it is fairly straightforward (since the epidemic process is a continuous-time Markov chain) to write code to simulate realisations of the final size of stochastic SIR epidemic process described in section *Model description* of the main article. By inspecting histograms of these simulated final sizes we find that the cut-off of 10% of population size (as stated in section *Model description* of the main text) is satisfactory for the networks we consider.

In plots depicting the results of simulations we present point estimates, for example the proportion of simulations that result in a final size greater than 10% of population size as an estimate of major outbreak probability, the mean final size amongst outbreaks with a final size greater than 10% of population size as an estimate of the relative final size of a major outbreak. We also present confidence intervals (CI) around these estimates in cases where those intervals are large enough to be visible on the scales used in the plots.

In simulations that are based on the configuration network construction imperfections of the network such as self-loops and multiple edges may arise. However, there are generally only a small number of such imperfections. In Table 1 imperfections are considered as a function of the total number of edges in the population for different population sizes for the configuration network with degree distribution $p_{d=0}^{10}$ with $p_d = c/(d+1)$, and $c = 0.331$ the normalization constant. As predicted by theory, the average number of imperfections is independent of population size while the fraction (as fraction of the total number of edges in the network) tends to zero with growing population size.

Population size	Number of edges	Number of imperfections	Fraction of imperfections
100	134	10.06	0.075
500	663	10.95	0.017
1000	1318	10.85	0.008
5000	6612	10.90	0.002

Table 1: The average number of self-loops and multiple edges for different population sizes for the configuration network with degree distribution $\{p_d\}_{d=0}^{10}$ with $p_d = c/(d+1)$, and $c = 0.331$ the normalization constant. The last column considers the average number of imperfections as fraction of the total number of edges in the network. The average is taken over 100 realizations of the configuration network for each population size.

References

- [1] F. Ball, D. Mollison, and G. Scalia-Tomba. Epidemics with two levels of mixing. *Ann. Appl. Probab.*, 7(1):46–89, 1997.
- [2] F. Ball, D. Sirl, and P. Trapman. Threshold behaviour and final outcome of an epidemic on a random network with household structure. *Adv Appl Probab*, 41:765–796, 2009.
- [3] F. Ball, D. Sirl, and P. Trapman. Analysis of a stochastic SIR epidemic on a random network incorporating household structure. *Math Biosci*, 224:53–73, 2010.
- [4] F. Ball, L. Pellis, and P. Trapman. Reproduction numbers for epidemic models with households and other social structures II: Comparisons and implications for vaccination. *Math. Biosci.*, 274:108–139, 2016.

- [5] F. Ball, T. Britton, K. Y. Leung, and D. Sirl. An SIR network epidemic model with preventive dropping of edges. *in preparation*, 2018.
- [6] N. G. Becker and K. Dietz. The effect of household distribution on transmission and control of highly infectious diseases. *Math. Biosci.*, 127:207–219, 1995.
- [7] B. Bollobás. *Random Graphs*. Academic Press, New York, 2001.
- [8] T. Britton, D. Juher, and J. Saldaña. A network epidemic model with preventive rewiring: comparative analysis of the initial phase. *Bull Math Biol*, 78:2427–2454, 2016.
- [9] Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data>, June 2014.