Social network-based distancing strategies to flatten the COVID 19 curve in a post-lockdown world

Preprint · April 2020

CITATIONS
0

READS
213

7 authors, including:

Per Block
University of Oxford
28 PUBLICATIONS **353** CITATIONS
SEE PROFILE

Marion Hoffman
ETH Zurich
2 PUBLICATIONS **1** CITATION
SEE PROFILE

Isabel J Raabe
University of Zurich
6 PUBLICATIONS **29** CITATIONS
SEE PROFILE

Jennifer Beam Dowd
University of Oxford
92 PUBLICATIONS **3,236** CITATIONS
SEE PROFILE

Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world

Authors

Per Block[1,*], Marion Hoffman[2], Isabel J. Raabe[3], Jennifer Beam Dowd[1], Charles Rahal[1], Ridhi Kashyap[1], Melinda C. Mills[1,*]

[1] Leverhulme Centre for Demographic Science & Department of Sociology, University of Oxford, UK.

[2] Department Humanities, Social and Political Sciences, ETH Zurich, Switzerland.

[3] Institute of Sociology, University of Zurich, Switzerland.

*Corresponding authors: per.block@sociology.ox.ac.uk and melinda.mills@nuffield.ox.ac.uk

Abstract

The COVID-19 pandemic highlights the importance of effective non-pharmaceutical public health interventions. While social distancing and isolation has been introduced widely, more moderate contact reduction policies could become desirable owing to adverse social, psychological, and economic consequences of a complete or near-complete lockdown. Adopting a novel social network approach, we evaluate the effectiveness of three targeted distancing strategies designed to 'keep the curve flat' and aid compliance in a post-lockdown world. We simulate stochastic infection curves that incorporate core elements from infection models, ideal-type social network models, and statistical relational event models. Our models demonstrate that while social distancing measures clearly do flatten the curve, strategic reduction of contact can strongly increase their efficiency, introducing the possibility of allowing some social contact while keeping risks low. Limiting interaction to a few repeated

contacts emerges as the most effective strategy. Maintaining similarity across contacts and the strengthening of communities via triadic strategies are also highly effective. This approach provides empirical evidence which adds nuanced policy advice for effective social distancing that can mitigate adverse consequences of social isolation.

Introduction

The non-pharmaceutical intervention of 'social distancing' represents a central policy in the battle to reduce the spread of COVID-19. The policy attempts to slow the speed of disease transmission by maintaining a physical distance between people and reducing social interactions *(1)*. The aim is to slow the growth rate of infections and avoid overburdening health-care systems, a strategy now widely known as 'flattening the curve' *(2)*. Social distancing measures include bans on public events, the closure of schools, universities and non-essential workplaces, the limiting public transportation, travel and movement restrictions, and public information campaigns urging citizens to voluntarily alter their private social interactions.

The majority of existing empirical studies on mitigating influenza pandemics focus on the effectiveness of different individual measures, such as travel restrictions, school closures or vaccines *(3,4)*. Few have considered interventions and the structure of social networks concurrently. When social networks are examined, it is generally in relation to vaccination strategies *(5)*, contact tracing, or analysing the spread of the virus through social networks *(6,7)*. Here we outline key **behavioural strategies for selective contact reduction that every individual and organisation can adopt** in order to maximise the benefits of limiting contact and social distancing. Applying insights from social network research and statistical network models, we demonstrate how the rate and extent of the spread of the virus differs by individuals' specific choices and organisational routines to change local contact network

configurations. We introduce and assess the impact of three distinct strategies: interrupting geographically and socio-demographically distant contacts, decreasing ties that bridge social clusters, and creating 'micro-communities' by repeatedly interacting with the same partners.

Since the timing of the development of a vaccine is uncertain, it is vital to develop effective non-pharmaceutical interventions that can be sustained in the medium to long term. Although current guidelines in many countries recommend or even mandate by law severe restrictions regarding social interaction at some point, some contact may be unavoidable, such as the provision of care. Furthermore, as time passes, issues related to the social, psychological, and economic cost — in addition to compliance fatigue — are likely to emerge, which requires the consideration and evaluation of different approaches to the practice of social distancing over the long-term. Once the first peak of the epidemic has passed and strict restrictions can be eased, such longer-term strategies are likely to be needed to avoid a resurgence of an unhindered second wave of infection.

Our approach recognises the cost of — and potential compliance fatigue with — complete isolation *(8).* Fully quarantining non-infected, psychologically vulnerable individuals over prolonged periods can have severe mental health consequences, and the strain of isolation can foster a rise in stress, negative emotions and domestic violence. Due to these costs, compliance with recommendations to *strategically reduce* contact is likely to be higher than compliance with *complete isolation*. Targeted recommendations for the strategic alteration of social contacts can mitigate psychological and physical harm as well as going a long way to contain of COVID-19 and, thus, keep the curve flat in the longer run.

The effectiveness of non-pharmaceutical and social distancing interventions

Scientific evidence on the effectiveness of non-pharmaceutical public health interventions has often relied on 'expert recommendations' rather than a strong scientific evidence base *(8,9).*

The most common and efficient interventions are the promotion of hand hygiene and respiratory etiquette *(10)*. Evidence regarding the effectiveness of masks and personal protective equipment for the general public is mixed *(11)*. Another common strategy — the isolation of infected individuals within healthcare facilities — is infeasible over the course of widespread epidemics such as COVID-19, leading to the necessary adoption of social isolation and distancing measures.

During previous outbreaks (e.g. SARS-CoV), social distancing measures such as workplace closures, limitation on public gatherings, and travel restrictions were implemented. The cancelling of public gatherings and long-distance travel restrictions appears to decrease rates of transmission and morbidity *(11)*. There is mixed evidence of the effectiveness of school closures on respiratory infections, possibly because of timing of school closures in the outbreak, or its effect only on school-aged children *(12)*.

There has been considerably less research on the effectiveness of types of social distancing measures beyond the interventions discussed above. Social separation for a long period may have adverse effects, such as loneliness and mental health, but also disrupts economic life (workplace), faith (religious worship), and family (visits of those in care, funeral attendance). Existing research has demonstrated that interventions are only effective and feasible when the public deems them acceptable *(11)*. For this reason, we adopt a novel approach that assesses the effectiveness of adaptions that rely on less confinement.

A network approach to flattening the curve

Flattening the (infection) curve represents a decrease in the number of infected individuals at the height of the epidemic and distributing the incidence of cases over a longer time span *(2)*. This is largely achieved by reducing the reproduction number (R), which is how many individuals are infected by each carrier. Reducing the reproduction number decreases the

growth rate of infected cases. Social distancing policies are therefore implicitly designed with the objective of limiting the *amount* of social contact between individuals. By introducing a social network approach, we propose that a decrease in the reproduction number can simultaneously be achieved by managing the *network structure* of interpersonal contact.

From a social network perspective, the infection curve is closely related to the concept of *network distance* (or *path lengths*) *(13)*, which indicates the number of network steps needed to connect two nodes. Popularised examples of *network distance* include the 'six degrees of separation' phenomenon *(14)*, which claims that any two people are connected through at most five acquaintances, the closely related 'six degrees of Kevin Bacon' which applies the concept to co-appearance in movies, and the infamous Erdös number: the number of "hops" needed to connect the author of a paper with the prolific late mathematician Paul Erdős.

The relation between infection curves and *network distance* can be illustrated with a simple network infection model as shown in Figure 1. Panels A and C depict two networks with different path lengths, each with one hypothetically infected COVID-19 seed node (purple square). At each time step, the disease spreads from infected nodes to every node to which they are connected; thus, in the first step the disease spreads from the seed node to its direct neighbours. In the second step, it spreads to their neighbours, who are at network distance 2 from the seed node, and so forth. Over time, the virus moves along network ties until all nodes are infected. The example shows that the network distance of a node from the infection source (indicated by node colour in Fig. 1, A and C) is identical to the number of time-steps until the virus reaches it. Thus, the distribution of distances to the source directly maps onto the curve of new infections (Fig. 1, B and D).

In our example, both networks have the same number of nodes (individuals) and edges (interactions); however, the network depicted in panel C has a much flatter infection curve than the network in panel A, even though eventually all nodes are infected in either case. This is

because the latter network has longer path lengths than the former one – or in other words, more network distance between the individuals (nodes) due to a different *structure* of interaction, despite the same absolute contact prevalence. Thus, when adopting a network perspective, flattening the curve is equivalent to increasing the network path from an infected individual to all others, which can be achieved by *restructuring* contact (next to the generally proposed *reduction* of contact). Consequently, one aim of social distancing should be increasing the *average network distance* between individuals by smartly manipulating the structure of interactions. The aim of this study is to provide insight on how to keep the COVID-19 curve flat while allowing some social interaction. Our illustration shows a viable path towards achieving our goal: we must devise interaction strategies that make real-life networks look more like network C, and less like network A.

Strategies to increase network distance

Following official social distancing advice, the most critical strategy to increase infection path lengths is to minimise social contact whenever possible. When enough social contact is interrupted, the network distance between healthy individuals and those with COVID-19 will become very long on average, regardless of which connections in particular are severed. However, as we demonstrate in this section, similar increases in path length can be obtained by strategically reducing contacts. Understanding which types of strategies are more efficient in increasing path lengths and flattening the curve can inform guidance to shift from short-term (complete lockdown) to long-term management of COVID-19 contagion processes.

The contact reduction strategies we propose are based on insights regarding how items flow through networks, such as diseases, memes, information, or ideas *(15, 16, 17, 18)*. Such spread is generally hampered when networks consist of densely connected groups with few connections in-between, such as among individuals who live in isolated villages scattered over

sparsely inhabited rural areas *(19)*. In contrast, contacts that bridge large distances are related to short paths and rapid spread, For example, when commuters travel between these isolated villages, network distances decrease substantially *(15, 19)*. Using this knowledge, we can avoid rapid contagion by encouraging social distancing strategies that increase clustering (such as staying within a village) and reduce network short-cuts (such as not commuting between villages) to reap the largest benefit of reducing social contact while keeping the limiting disease spread to a minimum. We propose three strategies that are all aimed at increasing network clustering and eliminating short-cuts.

While a more realistic example of the proposed strategies is simulated in the next section, we outline the underlying principles of the model in Figure 2. Panel A depicts a network in which densely connected communities are bridged by random, long-range ties; this type of network is commonly known as a *small world network (19)*. It is widely used in simulations, as it represents core features of real-world contact networks, in particular social clustering combined with short network distances, which makes it particularly useful for our illustration *(15)*. Panels A to D illustrate the successive, targeted contact reduction strategies, while the bar-graph depicts the distribution of distances of all individuals from one of the two highlighted infection sources.

Strategy 1: 'Birds of a feather' homophily strategy: Reduce geographic, organisational and socio-demographic difference to contact partners (A to B in Fig. 2)

To implement the first strategy, individuals need to pay attention to characteristics of their contact partners. Individuals tend to have contact with others which share common attributes, such as the neighbourhood they live in (geographical), the companies they work at (organisational), or that are of similar age (demographic) *(20,21,22)*. This is called 'homophily' and is a ubiquitous and well-established feature of social networks. The node colour in Figure

1 can represent any of these characteristics, such as neighbourhood of residence. Because we are mostly connected to similar others, contact with *dissimilar* individuals tends to bridge to more distant communities. Restricting one's contact to those most similar helps limit potential network bridges that are known to substantially reduce network path lengths. Panel B in Figure 1 shows the resulting network structure after this strategy of tie reduction has been implemented. The associated bar-graph illustrates that a substantial number of nodes are at a larger distance from the infection source following this network intervention. As indicated in the figure, this strategy promises success in case the characteristic takes on a variety of values for different individuals that promotes forming small communities. A simple split such as for example, along gender or ethnic lines does not promise measurable success, but will instead increase negative consequences of distancing measures.

This strategy is supported by epidemiological modelling which suggests that co-residence and mixing of individuals from different ages strongly increased the spread of infectious disease, such as COVID-19, in some countries *(23)*. Providing a concrete example of a behavioural recommendation for a post-lockdown world, if people only interact with others that live within 3 blocks (increase geographic similarity by constraining the area radius of gyration), more than 30 transmission events would be necessary for a disease to travel 100 blocks. Workplaces where many individuals need to come together can similarly implement routines to limit contact between groups that live in the same geographically distinct areas or come from different age-groups. The more similar a potential contact is, the closer he/she lives, and the more organisations that are shared (e.g., working in the same team; children in the same classroom), then the lower the comparative risk to keeping this contact.

Strategy 2: Strengthen community cohesion triadic strategy: Increased clustering among contact partners (B to C in Fig. 2)

For the second strategy, individuals must consider with whom their contact partners usually interact. A common feature of contact networks is 'triadic closure', referring to the fact that contact partners of an individual tend to be connected themselves *(21, 24, 25)*. Tie embedding in triads is a particularly useful topology for containing epidemic outbreaks. Consider a closed triad of individuals *i, j,* and *h*. When *i* infects *j* and *h*, the connection between *j* and *h* does not contribute to further disease spread: this contact is *redundant (26)*. When comparing networks with an identical number of connections, networks with more redundant ties tend to have longer path lengths. Accordingly, when removing contact to others, one should prioritize removing those ties that are not embedded in triads, since these ties generally decrease path lengths more. In practice, this means that contact with people who are not also connected to one's usual social contacts should be curtailed. Panel C in Figure 1 illustrates the resulting structure if ties that are not part of closed triads or 4-cycles are removed. In the ideal-type example presented, this intervention not only further reduces the network distance of many nodes from the infection sources, but also creates isolated communities or 'pods' that cannot be infected by the virus.

Strategy 3: Create 'micro-communities' strategy: Repeated contact to same others, rather than changing interaction partners (C to D in Fig. 2)

For the third strategy, individuals need to pay attention to their latest realised interactions and restrict their interactions these same people. This strategy reduces the number of contact partners rather than number of interactions, which is particularly important when contact is necessary for psychological well-being. This strategy of limiting contact to very few others with repeated interactions can be in the spirit of a social contract with others to create micro-communities to only interact within the same group delineated by common agreement. While

this requires coordination between interaction partners to only meet with the same designated others, the resulting micro-communities are very difficult for a virus to enter from the outside, or — importantly — if the infection is contracted by one contact, for the virus to spread further. Another implication of this strategy with respect to repeat interaction partners includes having social contacts that overlap across more than one contact group. For example, meeting colleagues outside of work for socializing will have less of an impact on the spread of the virus relative to a separate group of friends, since a potential infection path already exists. Or having tight and consistent networks of medical or community-based carers for those vulnerable to COVID-19 (elderly, pre-existing conditions) limits the transmission chain. Organisations can equally leverage this strategy in structuring work-routines by streamlining work-flows so that individuals rely on repeat contact to few others rather than workers dispersing throughout an organisation. Panel D in Figure 2 illustrates the resulting network structure.

Strategy 2 and 3 are similar in that they build on pre-existing network structures. However, their difference lies in determinants of individual interaction. Strategy 2 relies on a stable and established network structure of durable relations: who are my 'friends' (contact partners) and which pairs of my friends also know each other? Strategy 3 relies on preceding meetings in the immediate past – whom did I *meet* last? In this sense strategy 2 is easier to implement, since the rules regarding whom to meet do not change over the short term, while strategy 3 requires coordination of action.

Network simulation of disease spread

To this point, the strategies were illustrated on an intuitive but highly stylized model of epidemic spread. We demonstrate how our three contact strategies impact on infection curves using more formal stochastic infection models that incorporate core elements from infection

models, ideal-type network models and statistical relational event models. The baseline (null) model represents how infection would spread in the case of no social distancing at all.

First, as in classical disease modelling, individuals (actors) in our simulation can be in three states: susceptible, infected, and removed (recovered and no longer susceptible to infection, or deceased). Most actors begin in the susceptible state, while *q* random actors are in the infected state (assumed as one per thousand in our simulations). This can represent, for example, the post-lockdown scenario in which only few cases of COVID-19 remain in the population; however, variation of *q* might also be used to determine the levels at which quarantine could end and we can enter a post-lockdown world. During the simulation, susceptible actors can transition to the infected state by having contact with infected others (contact partners will be called *alters* from here on). A designated time after becoming infected, actors move to the removed state.

Second, as in many previous modelling efforts regarding dynamics of epidemics such as influenza, we do not assume homogeneous contact probabilities between all pairs of actors in an affected population but impose a network structure that limits contact opportunities between actors *(27, 28, 29)*. This network represents the usual contact people had in a pre-COVID-19 world. The networks we generate stochastically for our model follow fairly standard ideal-type network generating approaches. Representing place of residence, actors are assumed to have a fixed geographic location, represented by coordinates in a two-dimensional space. They are members of groups, such as households, institutions or workplaces, and have individual attributes, such as age, education, or income. Network ties are generated so that actors have some connections to geographically close alters, some ties to members of the same groups (representing e.g., co-workers), some ties to alters with similar attributes (e.g., similar age), and, finally, some ties to random alters in the population (e.g., haphazardly acquired friends not captured by the previous variables). The exact algorithms which define the networks are

described in the Methods section. The importance of different tie formation mechanisms, the total number of network ties of actors, the size of the population and so forth are varied in different simulation scenarios to ensure that results are independent of arbitrary assumptions about specific network features within the modelling framework. Descriptive statistics of the different scenarios employed in the simulations are presented in Table 1 and parameters to generate these are presented in Table S1 in the Appendix. The generated networks represent the structure of alters that an actor can possibly interact with. They represent the members of their so-called 'social circles' *(20, 21, 30)* with whom they interact with in normal, pre-COVID life (including family, friends, schoolmates or co-workers).

As the third component of the model, actors in the network interact at discrete times with alters with which they have a network connection, or in other words, they meet somebody from within their social circle or usual social contacts. This represents the actual contact people have in their lives. In case either interaction partner is infected but the other one is not, the disease is transmitted from the infected actor to the susceptible one during these interaction events.

However, in contrast to other modelling approaches, we do not assume that actors interact with random alters in their personal network but, rather, that they are purposeful actors who can make strategic choices with whom to interact. These strategic choices are at the core of our advice for policy interventions for individuals in attempts to increase the efficiency of social distancing. In our model, all choices are stochastic; strategies increase the likelihood of interacting with specific alters but are not deterministic. The exact formulation of with whom to interact follows a multinomial logit model to choose among the possible interaction partners as given by the network structure. This type of model has previously been used in network evolution models *(31)* and relational event models *(32, 33)*, as outlined in the Methods section.

As discussed above, our simulations explore three interaction strategies. First, in our 'birds of a feather' homophily strategy, actors choose to interact predominantly with others that are

similar to themselves based on a specified attribute used at the network generation stage. This could represent for instance interaction only with those within their age group or neighbourhood. Second, actors can adopt our 'strengthen community' or triadic strategy and chose to mostly interact with alters that have common connections in the underlying network, for example as common among family members. Third, adopting our 'repeated contact' strategy, actors can base their interaction choices on whom they have interacted with in their previous contacts, both as sender and receiver of an interaction. This might be, for instance, repeated care exchanges by the same individuals for a vulnerable or elderly neighbour or interactions between the same (two) households.

In each case, a separate statistical parameter in the multinomial model determines the probabilities of interaction partners based on the: (i) similarity of alters, (ii) number of common contacts the actor and the alter have; and, (iii) repeat interaction with one of the last *j* contact partners (see Methods). These three strategies are compared to a baseline case that mirrors the simple reduction of contact in which individuals have the same amount of interactions but choose randomly among their network contacts (a naïve contact reduction strategy) and a null model that represents contact patterns without any social distancing.

To make the comparison of interaction *strategies* independent of the arbitrary size of statistical parameters, we need to find a common metric that means parameter sizes are comparable in the extent to which they influence strategy decisions. To achieve this, we empirically calibrate parameters so that the average entropy in the probability distribution that represents the likelihood of different interaction choices is identical for all strategies (see Methods) *(34)*.

Results

We present the average outcome of one scenario of our simulation study in Figure 3. The *x*-axis represents time as measured in simulation steps and the *y*-axis represents the number of

individuals infected out of a total population of 1,000. Curves are averaged over 48 simulation runs. The first scenario in blue shows an interaction model in which there is no social distancing and actors interact at random. It essentially serves as a null model. The next four strategies all employ a 50% contact reduction compared to this null model in order to compare the different contact reduction strategies. The black line represents naïve social distancing in which actors reduce their contact in a random fashion. The golden line represents the first 'birds of a feather' homophily approach, where we plot the infection curve when individuals employ the homophily strategy for their remaining interactions. The green line models our second strategy of strengthening communities and represents the infection curve when remaining contacts are chosen according to the triadic strategy. Finally, the red line models our third strategy of repeated interaction contacts and shows how infections develop using the repetition strategy to choose interaction partners.

This simulation shows that all three of our strategies slow the spread of the virus compared to either no intervention or simple social distancing. The most effective choice is a strategic reduction of interaction with specific repeated contacts (our third strategy), with the homophily and triadic strategies roughly effective at the same levels. The results presented here are consistent across the different network examples that we used for our simulations. We find that in all examples, strategic social distancing strategies are more effective than random reduction, and that repeated social contacts with the same individuals is the most efficient strategy. Whether homophily or triadic closure have a larger effect depends on the specific case. Basic metrics of more examples of simulations under different conditions are found in Table 2.

Discussion and conclusion

In lieu of an effective vaccine for COVID-19, governments need to implement effective non-pharmaceutical policies to 'flatten the curve' in order to save lives and protect health systems

from becoming overloaded. Social and physical distancing have been suggested as one of the most prominent public health interventions against COVID-19. While a complete or near-complete lockdown has been adopted in many countries in order to reduce the first peak of hospital admissions, some form of contact reduction is likely to be necessary in the longer run, possibly until a vaccine is widely available. Adopting a social network approach, we model and test the effectiveness of three simulated strategies that can limit COVID-19 spread as a longer-term, post-lockdown strategy, while allowing to some extent social and economic life to converge towards pre-COVID levels. The proposed strategies are: (1) increasing similarity in contact partners (birds of a feather strategy), (2) strengthening community structure via triadic interaction; and, (3) repetition of contacts in relatively closed social networks.

This study provides governments and policymakers with much sought-after scientific evidence regarding the nuanced effects of different social distancing measures beyond the blanket terminology that has been introduced until now. Since public adherence is essential in what may be months of social distancing measures, this study provides more concrete and tailored advice to the public to empower them with more knowledge in order to design their own personal distancing strategies to generate safe social networks in the medium to long term, both in the private sphere, schools and the workplace.

We anticipate that this social network approach offers considerable new insights. By enabling individuals to differentiate between 'high-impact' and 'low-impact' contacts — and thus to strategically adjust their behaviour higher levels of compliance could be the result. Instead of asking individuals to completely self-isolate alone, the emphasis on similar, community-based, and repetitive contact measures could make the persistent implementation of these measures more palatable. These more nuanced measures will lighten the psychological burden on those who are likely to suffer the most from social isolation, such as individuals with depression or anxiety or those in difficult living arrangements. At the same time, these measures might assist

organisations in structuring contact in economic life so that more businesses can remain operational to some degree by streamlining and restructuring intra-organisational contact, or re-opening schools, whilst not necessarily increasing susceptibility to Covid-19 or other future pandemics,

A number of concrete behavioural guidelines for different contexts can be deduced from these strategies. For hospital and other workers, it means keeping the composition of shifts constant over prolonged amounts of time (i.e., repeating contact) and, where possible, distribute people into shifts based on, for example, residential proximity (i.e., homophily). In other organisations, mandating different starting and end-times of workdays and break times by organisational unit can keep contact in small groups and reduce it between them. For schools it means maintaining the similar composition of classrooms, distribution of children within their grades to residential proximity and varying break times. The strategies outlined above can also be applied to private situations in which contact is essential. For example, when providing private care to elderly relatives or neighbours, our models suggest that it is important that the same person provides this care, rather than different relatives or individuals taking turns. The person who provides such care should be the one with fewest bridging ties to other groups and the one who lives geographically closest to minimise transmission risk for all involved and the wider population. Similarly, repeated social meetings of individuals of similar ages that live alone carry a comparatively low risk. However, when in a household of five, each person engages in social meeting with their own sets of friends, many short cuts are being formed that are potentially connected to a very high risk of spreading the disease. Similarly, groups of children playing in the streets potentially create infection paths between many, otherwise unconnected families.

In conclusion, we show that simple behavioural rules can go a long way in ‘keeping the curve flat’. Once the most stringent lockdown measures can be eased, policymakers that give guidance might supplement their appeal to minimize contact with the request to make smart

and personalized choices in which contacts are contrived, functional and recreational. By following these simple strategies of homophilous interactions, strengthening communities and repeated contact with the same individuals, governments will be able to more effectively describe the impact of social distancing and the public more likely to feel they have the control and understand their logic and follow them.

Methods

Generation of stylised networks

The stylised binary networks x that represent interaction opportunities of individuals are generated as the sum of four sub-processes, representing a tie generated in any sub-process that will be transferred to the overall network. Jointly, the sub-processes create networks that have realistic values of local clustering, path-lengths, and homophily. All ties in the network are defined as undirected. The number of actors in the network is denoted by n.

The first sub-process represents tie formation based on geographic proximity *(35)*. First, all actors in the network are randomly placed into a two-dimensional square. Second, each actor draws the number of contacts which it forms in this subprocess $d_{geo,i}$ from a uniform distribution between $d_{geo,min}$ and $d_{geo,max}$; for example, if $d_{geo,min} = 10$ and $d_{geo,max} = 20$, every actor forms a random number of ties between 10 and 20 in this sub-process. Third, the user-defined density in geographic tie-formation g_{geo} defines the geographic proximity of contacts drawn, so that actor i randomly forms $d_{geo,i}$ ties among those $d_{geo,i}/g_{geo}$ that are closed in Euclidean distance from actor i. For example, if actor i is posed to form $d_{geo,i} = 12$ ties and $g_{geo} = 0.5$, the actor randomly choses 12 out of the 24 closest alters to form a tie to. Across all simulated networks we set $g_{geo} = 0.3$. Fourth, unilateral choices (where only i

selected j but not vice versa) are symmetrised so that a non-directed connection exists between the actors.

The second sub-process represents tie formation in organisational foci, e.g. workplaces *(36)*. First, each actor is randomly assigned to a group so that all groups have on average m members. Second, each actor forms ties at random to other members within the same groups with a probability of g_{groups}. For example, when $m = 10$ and $g_{groups} = 0.5$, a tie from each actor to every alter in the same group is formed with a probability of 50%. Third, unilateral ties are symmetrised as above.

The third sub-process represents tie-formation based on homophily, for example similarity in age or income *(22)*. First, each actor is assigned an individual attribute a_i between 0 and 100 with uniform probability (the scale of a_i cancels later in the model). Second, for each actor, the normalised similarity $sim_{i,j}$ to all alters j is calculated, which is the absolute difference between a_i and a_j for actor j, divided by 100 (the range of the variable). Third, each actor draws the number of contacts it forms in this subprocess $d_{homo,i}$ from a uniform distribution between $d_{homo,min}$ and $d_{homo,max}$. Fourth, each actor creates $d_{homo,i}$ ties to alters j in the networks with a probability that is proportional to $(sim_{ij})^w$, where w regulates the extent to which individuals prefer more similar others. Across all reported simulations, we set $w = 2$. Fifth, unilateral ties are symmetrised as above.

The fourth sub-process represents haphazard ties that are not captured by any of the above processes. Here simply z ties per actor are created with respect to randomly chosen alters.

Definition of simulation model

Let the binary network x represent the underlying social ties between n individuals, labelled from 1 to n. Each node i is characterised by an attribute a_i such as its age.

Our model aims to reproduce the process of individuals interacting with some of their social connections. These interactions occur in a sequence, using mechanics similar to the standard Dynamic Actor-Oriented Model *(33)* which was designed for relational events. More specifically, our model builds upon the following assumptions:

1. At each step of the process, one individual is picked at random and initiates an interaction with a probability π, representing the percentage of contacts thy maintain.
2. An actor initiating an interaction can only pick one interaction partner. Only neighbours in the underlying network x can be chosen. The decision to interact is unilateral and depends on characteristics of the two persons.
3. A contagious individual infects a healthy person as soon as they interact.
4. All individuals fully recover from the disease at a fixed rate. Once they recover, they can no longer be infected.
5. The process ends once there is no longer anyone contagious.

Let N_i be the set of neighbours j of a given individual i in the network x. We define for each step t of the process, $L_i(j,t)$ as the number of prior interactions between i and a neighbour j, within the last K interactions of i. In our simulations, the number K was arbitrarily set to 2 but could be further adjusted.

For each neighbour $j \in N_i$, the value $s(i,j)$ represents the statistic driving the strategical choice of i to pick j. Specifically, it can be defined in three different ways depending on whether the homophily, the triadic, or the repetition strategy is chosen. The statistic $s_{homophily}$ accounts for the level of homophily between i and j, and $s_{triadic}$ corresponds to the number of neighbours they share, where $s_{repetition}$ is the count of previous interactions within the last K contacts of i. In practise, these statistics are calculated as:

$$s_{homophily}(i,j) = \frac{|a_i - a_j|}{range(a)}$$

$$s_{triadic}(i,j) = \sum_{k=1}^{n} x_{i,k}x_{j,k}$$

$$s_{repetition}(i,j) = L_i(j,t)$$

The probability for i to pick j is defined as a multinomial choice probability *(37)*, following the logic of the DyNAM *(33)* and previous stochastic network models *(31)*. The intuition behind this distribution is that each potential partner in N_i is assigned an objective function value, and choosing a partner is based on these values. Mathematically, the objective function is an exponentiated linear function of the statistic $s(i,j)$, weighted by a parameter β. We further assume that individuals can reduce a certain percentage (π) of their interactions. Considering the probability (π) of initiating an interaction in the first place, the relevant probability distribution becomes:

$$p(i \to j \mid \pi, \beta) = \frac{\pi \exp(\beta * s(i,j))}{\sum_{j' \in N_i} \exp(\beta * s(i,j'))}.$$

These probabilities can be loosely interpreted in terms of log-odd ratios, similarly to logit models. Given two neighbours $j_1$and j_2 for which the statistic s increases of one unit (i.e., $s(i,j_2) = s(i,j_1) + 1$), the following log ratio simplifies to:

$$\log \frac{p(i \to j_2 \mid \pi, \beta)}{p(i \to j_1 \mid \pi, \beta)} = \beta.$$

For example, if we use $s = s_{repetition}$ and $\beta_{repetition} = \log(2)$, the probability of picking one neighbour present in the last contacts of i is twice as high as picking another neighbour who is not. The simulation of the model steps for different parametrizations were implemented in R and are available in the Supplementary Material.

Calibration of model parameters

The strategy of picking a neighbour at random corresponds to the model without any statistic s, reducing the probability distribution to a uniform one. For the three other strategies, the parameters $\beta_{homophily}$, $\beta_{triadic}$, and $\beta_{repetition}$ are adjusted to keep the models comparable.

To this end, we use the measure of explained variation for dynamic network models devised by Snijders *(34)*. This measure builds upon the Shannon entropy and can be applied to our model to assess the degree of certainty in the choices individuals make. For a given individual i at a step t, this measure is defined as:

$$r_H(i, t \mid \pi, \beta) = 1 + \frac{\sum_{j \in N_i} p(i \to j \mid \pi, \beta) \log_2(p(i \to j \mid \pi, \beta))}{\log_2(|N_i|)}.$$

Intuitively, this measure equals 0 in the case of the random strategy where the probability of picking any is neighbour identical. It increases whenever some outcomes are favoured over others, and equals 1 if one outcome has all of the probability mass.

Since the model assumes all individuals are equally likely to initiate interactions, we can average this measure over all actors. Moreover, in the case of the repetition strategy, the measure is time dependent. In that case, we use its expected value over the whole process. We finally use the following aggregated measure in order to evaluate the certainty of outcomes of a specific strategy:

$$R_H(\pi, \beta) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[r_H(i, t)].$$

For this article, we first fix the parameter $\beta_{repetition}$ at a value of 2.5, and calculate an estimated value $\widehat{R_H}(\pi, \beta_{repetition})$ of this measure. This experience-based parameter choice results in an associated R_H value between 0.3 and 0.5 in the different scenario, which is realistic in terms of size (see definition above). To compare this model to others, we then define the parameters $\beta_{homophily}$ and $\beta_{triadic}$ that verify

$$\widehat{R_H}(\pi, \beta_{repetition}) = R_H(\pi, \beta_{homophily}) = R_H(\pi, \beta_{triadic})$$

using a standard optimisation algorithm. The R code associated to these calculations can be found in the Supplementary Material.

References

1. Glass, R. J., Glass, L. M., Beyeler, W. E. & Min, H. J. Targeted social distancing design for pandemic influenza. *Emerg. Infect. Dis.* (2006).
2. Roberts, S. Flattening the Cornavirus Curve. *New York Times* (2020).
3. Ferguson, N. M. *et al.* Strategies for mitigating an influenza pandemic. *Nature* **442**, 448–452 (2006).
4. Germann, T. C., Kadau, K., Longini, I. M. & Macken, C. A. Mitigation strategies for pandemic influenza in the United States. *Proc. Natl. Acad. Sci.* **103**, 5935–5940 (2006).
5. Ventresca, M. & Aleman, D. Evaluation of strategies to mitigate contagion spread using social network characteristics. *Soc. Networks* **35**, 75–88 (2013).
6. Sun, K., Chen, J. & Viboud, C. Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study. *Lancet Digit. Heal.* **2**, e201–e208 (2020).
7. Wu, Z. & McGoogan, J. M. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China. *JAMA* (2020).
8. Morse, S. S., Garwin, R. L. & Olsiewski, P. J. PUBLIC HEALTH: Next Flu Pandemic: What to Do Until the Vaccine Arrives? *Science (80).* **314**, 929–929 (2006).
9. Nonpharmaceutical Interventions for Pandemic Influenza, National and Community Measures. *Emerg. Infect. Dis.* **12**, 88–94 (2012).

10. Bloomfield, S. F., Aiello, A. E., Cookson, B., O'Boyle, C. & Larson, E. L. The effectiveness of hand hygiene procedures in reducing the risks of infections in home and community settings including handwashing and alcohol-based hand sanitizers. *Am. J. Infect. Control* **35**, S27–S64 (2007).
11. Aledort, J. E., Lurie, N., Wasserman, J. & Bozzette, S. A. Non-pharmaceutical public health interventions for pandemic influenza: an evaluation of the evidence base. *BMC Public Health* **7**, 208 (2007).
12. Jackson, C., Vynnycky, E., Hawker, J., Olowokure, B. & Mangtani, P. School closures and influenza: systematic review of epidemiological studies. *BMJ Open* **3**, e002149 (2013).
13. Wasserman, S. & Faust, K. *Social network analysis : methods and applications*. (Cambridge University Press, 1994).
14. Milgram, S. The small world problem. *Psychol. Today* **2**, 60–67 (1967).
15. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
16. Podolny, J. M. Networks as the pipes and prisms of the market. *Am. J. Sociol.* (2001).
17. Borgatti, S. P. Centrality and network flow. *Soc. Networks* (2005).
18. Centola, D. The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010).
19. Watts, D. J. Networks, Dynamics, and the Small-World Phenomenon. *Am. J. Sociol.* **105**, 493–527 (1999).
20. Feld, S. The focused organization of social ties. *Am. J. Sociol.* **86**, 1015–1035 (1981).
21. Rivera, M. T., Soderstrom, S. B. & Uzzi, B. Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms. *Annu. Rev. Sociol.* **36**, 91–115 (2010).

22. McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a Feather: Homophily in Social Networks. *Annu. Rev. Sociol.* **27**, 415–444 (2001).

23. Pellis, L., Cauchemez, S., Ferguson, N. M. & Fraser, C. Systematic selection between age and household structure for models aimed at emerging epidemic predictions. *Nat. Commun.* **11**, 906 (2020).

24. Granovetter, M. S. The Strength of Weak Ties. *Am. J. Sociol.* **78**, 1360–1380 (1973).

25. Goodreau, S. M., Kitts, J. A. & Morris, M. Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography* **46**, 103–125 (2009).

26. Burt, R. S. *Structural Holes: The Social Structure of Competition.* (Harvard University Press, 1992).

27. Newman, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E* (2002).

28. Halloran, M. E. *et al.* Modeling targeted layered containment of an influenza pandemic in the United States. *Proc. Natl. Acad. Sci.* (2008).

29. Salathé, M. *et al.* A high-resolution human contact network for infectious disease transmission. *Proc. Natl. Acad. Sci.* (2010).

30. Block, P. Network evolution and social situations. *Sociol. Sci.* (2018).

31. Snijders, T. A. B. The Statistical Evaluation of Social Network Dynamics. *Sociol. Methodol.* **31**, 1–33 (2001).

32. Butts, C. T. A Relational Event Framework for Social Action. *Sociol. Methodol.* **38**, 155–200 (2008).

33. Stadtfeld, C. & Block, P. Interactions, Actors, and Time: Dynamic Network Actor Models for Relational Events. *Sociol. Sci.* **4**, 318–352 (2017).

34. Snijders, T. A. B. Explained Variation in Dynamic Network Models. *Math. Sci. hum.* **42**, 5–15 (2004).

35 1. Hamill, L. & Gilbert, N. Social circles: A simple structure for agent-based social network models. *JASSS* (2009).

36. Hébert-Dufresne, L. & Althouse, B. M. Complex dynamics of synergistic coinfections on realistically clustered networks. *Proc. Natl. Acad. Sci.* (2015).

37. McFadden, D. L. Conditional Logit Analysis of Qualitative Choice Behavior. *Frontiers in Econometrics* (1974).

Author contributions

PB, MH, & IJR conceptualised the study; PB & MH contributed methodology and implementation; PB, MH, IJR, & MCM wrote initial manuscript and provided visualisation; all authors reviewed and edited manuscript.

Competing interests

The authors declare no competing interests.

Data and materials availability

All functions and an example script are attached as Supplementary Material to this article.

Tables and Figures

Table 1: Characteristics of the networks created under different scenarios, averaged over 48 simulations in each case. *Notes:* n: number of actors; deg.: average degree / number of connections per actor; clus.: clustering coefficient / proportion of closed triads over possibly closed triads; av. *P*ath: average network distance between pairs of nodes; dia.: diameter / maximum distance in between nodes in the network; hom.: average similarity of interaction partners divided by average similarity among all actors.

Scenario	n	deg.	clus.	av. path	dia.	hom.
1: baseline scenario	1000	38.4	0.11	2.23	3.0	1.08
2: higher degree	1000	75.9	0.14	1.93	3.0	1.08
3: lower degree	1000	19.4	0.09	2.69	4.0	1.08
4: no groups	1000	55.4	0.16	2.07	3.0	1.07
5: no geography	1000	40.2	0.26	2.24	3.0	1.09
6: random net.	1000	62.0	0.06	1.96	3.0	1.00
7: geography	1000	53.9	0.30	2.57	4.0	1.00
8: 2000 actors	2000	38.8	0.09	2.49	3.3	1.08

Table 2: Characteristics of average infection curves for different strategies. *Notes*: all entries denoting averaged results of simulations are relative to the null model of no contact reduction (blue line in Fig. 3). delay: delay of the peak of the infection curve compared to the null model; peak: height of the peak of the infection curve compared to the null model; inf.: proportion of the population infected compared to the null model.

Scenario	Strategy	delay	peak	inf.
1: baseline scenario	random	2.32	0.39	89%
	triads	3.61	0.26	82%
	homophily	3.37	0.24	87%
	repetition	3.13	0.10	39%
2: higher degree	random	2.36	0.38	88%
	triads	2.66	0.35	88%
	homophily	2.51	0.36	89%
	repetition	1.61	0.12	34%
3: lower degree	random	2.36	0.38	89%
	triads	3.70	0.05	31%
	homophily	4.62	0.09	57%
	repetition	0.73	0.01	2%
4: no groups	random	2.36	0.40	90%
	triads	3.64	0.25	85%
	homophily	2.81	0.31	89%
	repetition	2.07	0.11	36%
5: no geography	random	2.30	0.38	89%
	triads	0.68	0.01	2%
	homophily	3.14	0.24	86%
	repetition	2.21	0.08	31%
6: random net.	random	2.27	0.39	89%
	triads	2.45	0.34	85%
	homophily	2.46	0.35	89%
	repetition	1.72	0.11	34%
7: geography	random	2.34	0.38	90%
	triads	3.70	0.24	83%
	homophily	2.79	0.33	88%
	repetition	2.06	0.11	38%
8: 2000 actors	random	2.39	0.37	89%
	triads	3.31	0.24	81%
	homophily	3.03	0.24	87%
	repetition	3.78	0.14	60%

Figure 1: Two example networks A and C with the same number of nodes (individuals) and ties (social interactions) but different structures that imply different infection curves B and D. Bold ties highlight the shortest infection path from the infection source to the last infected individual in the respective networks. Network node colour indicates at which step a node is infected and maps onto colours of histogram bars.

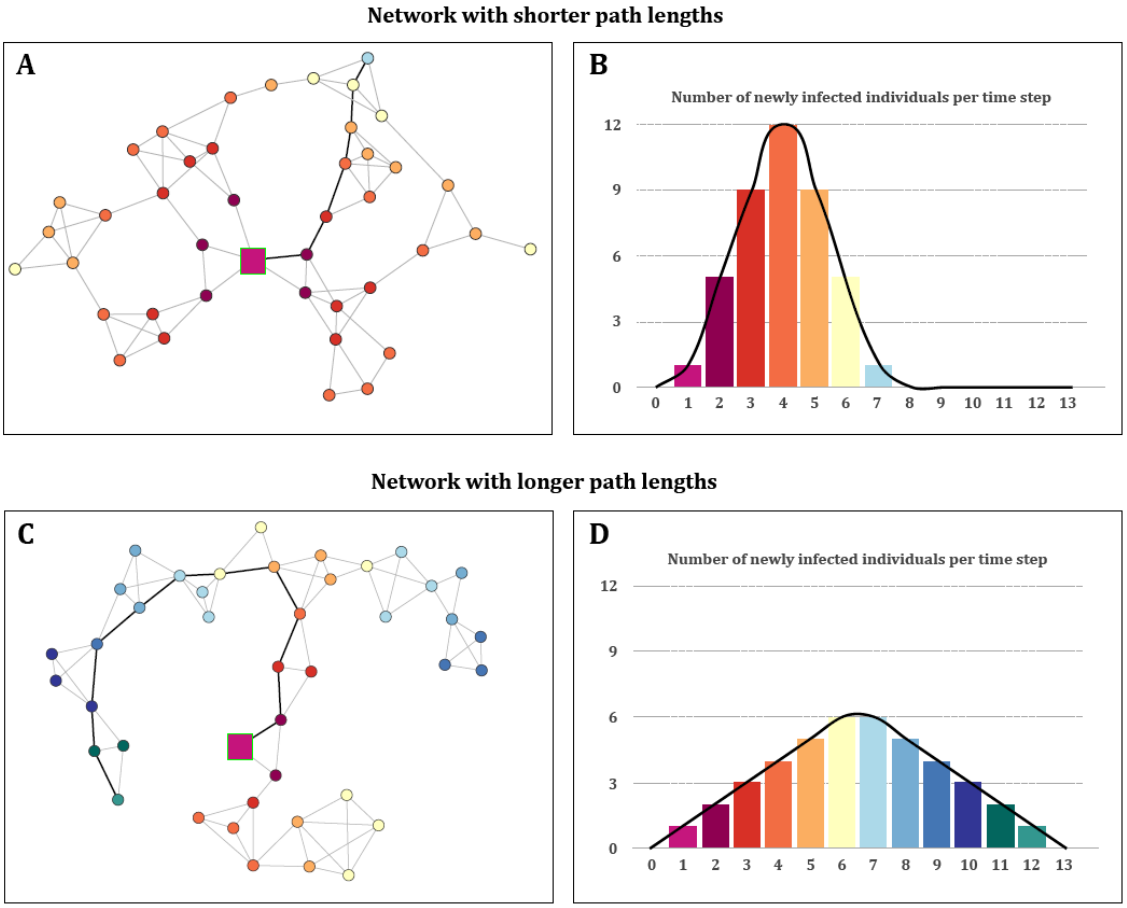


Figure 2: Example networks that result from the successive tie reduction strategies: A: initial small world network; B: removing geographically distant ties; C: removing non-embedded ties; D: repeat rather than extend contact; and E: network distances from the infection sources for examples.

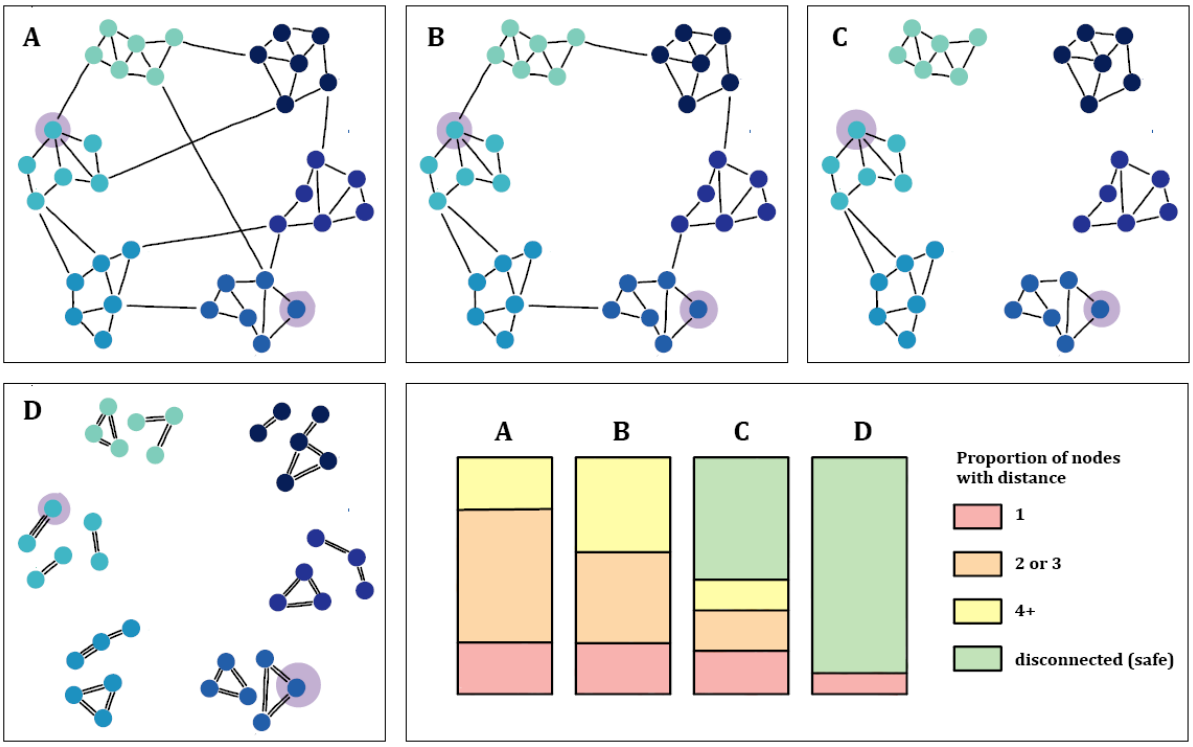
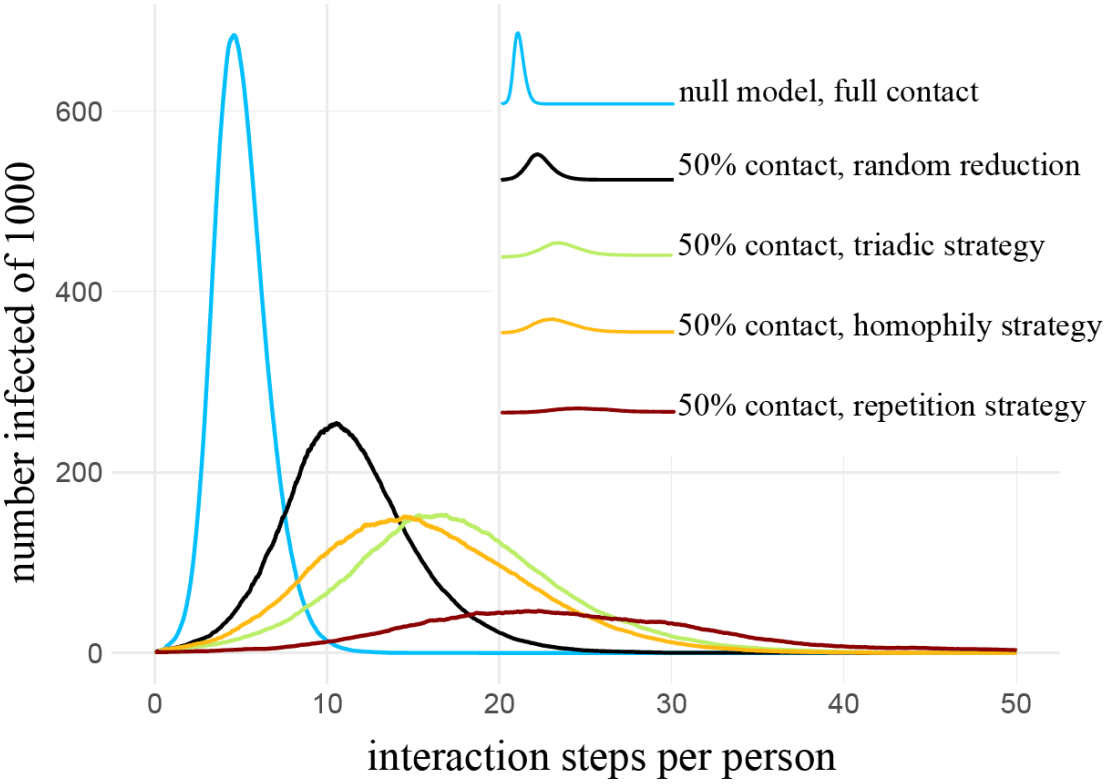


Figure 3: Average infection curves comparing 4 contact reduction strategies to the null model. Underlying network structure includes 1000 actors and is denoted 1: baseline scenario in the accompanying tables.



Supplementary Information to

Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world

Per Block[1,*], Marion Hoffman[2], Isabel J. Raabe[3], Jennifer Beam Dowd[1], Charles Rahal[1], Ridhi Kashyap[1], Melinda C. Mills[1,*]

[1] Leverhulme Centre for Demographic Science & Department of Sociology, University of Oxford, UK.

[2] Department Humanities, Social and Political Sciences, ETH Zurich, Switzerland.

[3] Institute of Sociology, University of Zurich, Switzerland.

*Corresponding authors: per.block@sociology.ox.ac.uk and melinda.mills@nuffield.ox.ac.uk

Content

Table S1

Scenario	nActors	d_geo_min	d_geo_max	g_geo	m_groups	g_groups	d_hom_min	d_hom_max	w_hom.	z_random
1: baseline scenario	1000	4	12	0.3	8	0.9	4	12	2	0.5
2: higher degree	1000	8	24	0.3	16	0.9	8	24	2	1
3: lower degree	1000	2	6	0.3	4	0.9	2	6	2	0.25
4: no groups	1000	10	30	0.3	0	NA	5	15	2	0.5
5: no geography	1000	0	0	NA	20	0.9	5	15	2	0.5
6: random net.	1000	0	0	NA	0	NA	0	0	NA	32
7: geography	1000	15	45	0.3	0	NA	0	0	NA	0.5
8: 2000 actors	2000	4	12	0.3	8	0.9	4	12	2	0.5

Notes: parameters used in the algorithm described in Methods and provided in the associated R script to generate underlying network structures

R-script “functions_SN_covid_spread.R”

```
##########################################################
### Simulation of social interactions with infections ###
### Supplementary to: "Efficient social distancing    ###
### strategies to reduce the spread of COVID"         ###
### Authors: Marion Hoffman, Per Block                ###
##########################################################


############ PART 1 ################
#
# generation of social networks
#
####################################
#                                  #
#  Make some networks that look    #
#  like real ones                  #
#                                  #
#  - distribute people in 2D space #
#  - make people have ties with    #
#  others very close               #
#  - add a few random ties         #
#  - add another var and add homo  #
#  philous ties                    #
#  - give people random group      #
#  membership                      #
#  - connect people within groups  #
#                                  #
####################################


socialCirclesNet <- function(nActors = 1000,
                             groupSize = 0, densityInGroups = 0.9,
                             minContactsGeo = 0, multMaxContactsGeo = 3, densityGeo
= 0.3,
                             randomTiesPerPerson = 0,
                             minContactsHomo   =   0,   multMaxContactsHomo   =   3,
homoStrength = 5,
                             orderAndUnique = T){

  ##### generate groups in which people are highly connected #####
  if(groupSize>0){
    nGroups <- round(nActors/groupSize)
    membership <- sample(1:nGroups, nActors, replace = T)
    group_edges    <-    which((((outer(membership,    membership,    "==")*1)    *
(runif(nActors^2) < densityInGroups)) == 1), arr.ind = T)
  }

  ##### generate geo based edges #####
  # Distribute people on 2D space

  geo_x <- runif(nActors) * 100
  geo_y <- runif(nActors) * 100

  if(minContactsGeo > 0){
    # draw number of contacts that everybody gets on geo

    geo_friends   <-   floor(runif(nActors,   min   =   minContactsGeo,   max   =
(minContactsGeo*multMaxContactsGeo + 1)))

    geo_data <- data.frame(x = geo_x, y = geo_y, altN = geo_friends)

    geo_edges <- lapply(1:nActors, function(i) geo.to.net(geo_data, i, densityGeo))
  }

  ##### generate homophilous edges #####

  coVar <- runif(nActors)

  if(minContactsHomo > 0){

    # draw number of contacts that everybody gets on homo
```

```
    homo_friends <- floor(runif(nActors, min = minContactsHomo, max =
(minContactsHomo*multMaxContactsHomo + 1)))

    homo_data <- data.frame(coVar = coVar, altN = homo_friends)

    homo_edges <- lapply(1:nActors, function(i) homo.to.net(homo_data, i,
homoStrength))
  }

  if(randomTiesPerPerson > 0){
    ##### generate random long range ties #####

    random_edges <- cbind(sample(1:nActors, round(nActors * randomTiesPerPerson),
replace = T),
                          sample(1:nActors, round(nActors * randomTiesPerPerson),
replace = T))
  }

  all_edges <- c(0,0)

  if(groupSize > 0){
    all_edges <- rbind(all_edges, group_edges)
  }

  if(randomTiesPerPerson > 0){
    all_edges <- rbind(all_edges, random_edges)
  }

  if(minContactsHomo > 0){
    all_edges <- rbind(all_edges, Reduce(rbind, homo_edges))
  }

  if(minContactsGeo > 0){
    all_edges <- rbind(all_edges, Reduce(rbind, geo_edges))
  }

  all_edges <- all_edges[-1,]

  row.names(all_edges) <- NULL
  colnames(all_edges) <- NULL

  if(orderAndUnique){
    all_edges <- all_edges[(all_edges[,1] != all_edges[,2]),]
    all_edges <- t(apply(all_edges, 1, sort))
    all_edges <- all_edges[order(all_edges[,1], all_edges[,2]),]
    all_edges <- unique(all_edges)
  }

  return(list(all_edges, coVar, geo_x))
}

##### functions used to make these networks #####

# function to take friends from geo data

geo.to.net <- function(dataframe, i, densityGeo){
  distanceToI <- sqrt((dataframe$x[i] - dataframe$x)^2 + (dataframe$y[i] -
dataframe$y)^2)

  maxDist <- sort(distanceToI)[dataframe$altN[i]/densityGeo + 1]

  alters <- which(distanceToI <= maxDist)[!which(distanceToI <= maxDist) %in% i]

  alters <- sample(alters, dataframe$altN[i], replace = F)

  return(cbind(i, alters))
}


# function to take friends from homo data

homo.to.net <- function(dataframe, i, homoStrength){

  simToI <- (1 - abs(dataframe$coVar[i] - dataframe$coVar))^homoStrength

  simToI[i] <- 0
```

```
  sampleProb <- simToI / sum(simToI)

  alters <- sample(1:nrow(dataframe), dataframe$altN[i], replace = F, prob =
sampleProb)

  return(cbind(i, alters))
}


############ PART 2 ################
#
# functions used as statistics in
# simulations
#
####################################

# function that returns the number of triads an edge is embedded in
calculate_triads <- function(graph,network){

  cocit <- cocitation(graph)
  triads <- cocit * network
  return(triads)
}

# function that returns the homophily of edges
calculate_homophilies <- function(network, attributes){

  range <- max(attributes) - min(attributes)

  sim <- 1 - (abs(outer(attributes, attributes, "-")) / range)

  homophilies <- sim * network

  return(homophilies)
}


# pre-calculation of the link functions for the triad and homophily effects
calculate_fixed_probas <- function(network, triads, homophilies, param_triad,
param_homophily) {

  probas <- param_triad*triads + param_homophily*homophilies
  probas <- exp(probas)
  probas[network == 0] <- 0
  return(probas)

}


############ PART 3 ################
#
# function to simulate infection
# spread
#
####################################


# Function to simulate contacts/infections
simulate_infection_curve <- function(network,
                                     triads = NULL,
                                     homophilies = NULL,
                                     percentage_keepcontact = 1,
                                     param_repetition,
                                     param_triad,
                                     param_homophily,
                                     window_repetition = 5,
                                     window_recovery = Inf,
                                     starting_nodes,
                                     burnin = NULL,
                                     lastcontacts = NULL,
                                     seed = 1){

  set.seed(seed)

  n <- nrow(network)
  if(is.null(triads)) triads <- sparseMatrix(i=1:n,j=1:n)
  if(is.null(homophilies)) homophilies <- sparseMatrix(i=1:n,j=1:n)
```

```
  # 1 calculate probabilities beforehand, possibly start keeping track of last
contacts
  probas_fixed <- calculate_fixed_probas(network, triads, homophilies, param_triad,
param_homophily)
  probas <- probas_fixed / matrix(rep(rowSums(probas_fixed),n),nrow=n,ncol=n)
  if(param_repetition != 0 && is.null(lastcontacts)) lastcontacts <- vector(mode =
"list", length = n)
  if(is.null(burnin)) burnin <- 0

  contacts <- list()
  infections <- list()
  step <- 1
  all_infected <- FALSE
  infected <- rep(-1,n)
  infected[starting_nodes] <- 0
  cpt_burnin <- 0

  while(!all_infected){

    keepcontact <- runif(1) <= percentage_keepcontact

    if(keepcontact) {

      # 2 pick a random node
      i <- sample(1:n,1)

      # 3 pick a neighbor
      neighbors <- which(network[i,] > 0)
      ps <- probas[i,neighbors]
      if(length(neighbors) > 1){
        j <- sample(neighbors, prob = ps, size=1)
      } else if(length(neighbors) == 1){
        j <- neighbors
      } else {
        next
      }
      contacts[[step]] <- c(i,j)

      # 4 infect and update the recovery
      if(cpt_burnin >= burnin) {
        newinfected <- infected
        if(infected[i] >= 0 && infected[j] == -1) newinfected[j] <- 0
        if(infected[i] == -1 && infected[j] >= 0) newinfected[i] <- 0
        if(infected[i] >= 0) newinfected[i] <- infected[i] + 1
        if(infected[j] >= 0) newinfected[j] <- infected[j] + 1
        infected <- newinfected
        infections[[step]] <- sum(infected >= 0)
      }

      # 5 update recovery
      if(cpt_burnin >= burnin) {
        if(infected[i] == window_recovery) infected[i] <- -2
        if(infected[j] == window_recovery) infected[j] <- -2
      }

      # 6 update network of repetition
      if(param_repetition != 0) {
        lci <- lastcontacts[[i]]
        lci <- c(j,lci)
        lci <- lci[1:min(window_repetition,length(lci))]
        lastcontacts[[i]] <- lci

        lcj <- lastcontacts[[j]]
        lcj <- c(i,lcj)
        lcj <- lcj[1:min(window_repetition,length(lcj))]
        lastcontacts[[j]] <- lcj
      }

      # 7 update probas
      if(param_repetition != 0) {
        ps <- probas_fixed[i,neighbors]
        t <- table(lci)
        newterms <- t[match(neighbors, names(t))]
        newterms[is.na(newterms)] <- 0
        ps <- ps * exp(param_repetition*newterms)
        probas[i,neighbors] <- as.vector(ps / sum(ps))
```

```
        neighbors <- which(network[j,] > 0)
        ps <- probas_fixed[j,neighbors]
        t <- table(lcj)
        newterms <- t[match(neighbors, names(t))]
        newterms[is.na(newterms)] <- 0
        ps <- ps * exp(param_repetition*newterms)
        probas[j,neighbors] <- as.vector(ps / sum(ps))
      }

    } else {

      # 4 update the recovery
      if(cpt_burnin >= burnin) {
        newinfected <- infected
        if(infected[i] >= 0) newinfected[i] <- infected[i] + 1
        if(infected[j] >= 0) newinfected[j] <- infected[j] + 1
        infected <- newinfected
        infections[[step]] <- sum(infected >= 0)
      }

      # 5 update recovery
      if(cpt_burnin >= burnin) {
        if(infected[i] == window_recovery) infected[i] <- -2
        if(infected[j] == window_recovery) infected[j] <- -2
      }
    }

    # check if half or all network is infected
    all_infected <- (sum(infected < 0) == n)
    step <- step + 1
    cpt_burnin <- cpt_burnin+1
  }

  return(list(contacts = contacts,
              infections = infections,
              number_steps_max = step-1,
              infected = infected))

}

# Umbrella function
simulate <- function(network,
                     triads,
                     homophilies,
                     percentage_keepcontact,
                     params,
                     window_recovery,
                     starting_nodes,
                     burnin,
                     lastcontacts,
                     seed){
  res <- simulate_infection_curve(network,
                                  triads,
                                  homophilies,
                                  percentage_keepcontact,
                                  param_repetition = params[3],
                                  param_triad = params[1],
                                  param_homophily = params[2],
                                  window_repetition = params[4],
                                  window_recovery,
                                  starting_nodes,
                                  burnin,
                                  lastcontacts,
                                  seed)
  return(res)
}


############ PART 4 ################
#
# functions related to entropy and
# estaimting proper parameter sizes
#
#####################################
```

```
# Calculate entropy (R_H) with no time-dependent statistics
calculate_entropy_fixed <- function(network,
                                    triads,
                                    homophilies,
                                    param_triad,
                                    param_homophily) {

  n <- nrow(network)
  degrees <- rowSums(network)
  isolates <- degrees == 0
  if(is.null(triads)) triads <- sparseMatrix(i=1:n,j=1:n)
  if(is.null(homophilies)) homophilies <- sparseMatrix(i=1:n,j=1:n)

  probas <- calculate_fixed_probas(network, triads, homophilies, param_triad,
param_homophily)
  probas <- probas / matrix(rep(rowSums(probas),n),nrow=n,ncol=n)
  fakeprobas <- probas
  fakeprobas[fakeprobas == 0] <- 1
  fakedegrees <- degrees
  fakedegrees[fakedegrees == 1] <- 2

  allRH <- 1 + rowSums(probas[!isolates,] * log2(fakeprobas[!isolates,])) /
log2(fakedegrees[!isolates])
  RH <- mean(allRH)
  return(RH)
}

# Calculate entropy (R_H) with time-dependent statistics (repetition)
calculate_entropy_timedependent <- function(network,
                                            triads,
                                            homophilies,
                                            param_triad,
                                            param_homophily,
                                            param_repetition,
                                            window_repetition,
                                            num_simulations,
                                            burnin,
                                            thining,
                                            seed) {

  set.seed(seed)

  n <- nrow(network)
  degrees <- rowSums(network)
  isolates <- degrees == 0

  if(is.null(triads)) triads <- sparseMatrix(i=1:n,j=1:n)
  if(is.null(homophilies)) homophilies <- sparseMatrix(i=1:n,j=1:n)
  probas_fixed <- calculate_fixed_probas(network, triads, homophilies, param_triad,
param_homophily)
  probas <- probas_fixed / matrix(rep(rowSums(probas_fixed),n),nrow=n,ncol=n)

  # initiate the last contacts
  lastcontacts <- vector(mode = "list", length = n)
  for(i in 1:n){
    neighbors <- which(network[i,] > 0)
    if(length(neighbors) > 1){
      lci <- c()
      for(k in 1:window_repetition){
        j <- sample(neighbors, size=1)
        lci <- c(lci,j)
      }
      lastcontacts[[i]] <- lci
    } else if(length(neighbors) == 1){
      lci <- rep(neighbors, window_repetition)
      lastcontacts[[i]] <- lci
    }
    ps <- probas_fixed[i,neighbors]
    t <- table(lci)
    newterms <- t[match(neighbors, names(t))]
    newterms[is.na(newterms)] <- 0
    ps <- ps * exp(param_repetition*newterms)
    probas[i,neighbors] <- as.vector(ps / sum(ps))
  }


  # simulate some steps and store all probas
```

```
  step <- 1
  cpt_burnin <- 0
  cpt_thining <- 0
  averageRH <- 0
  ntotal <- 0
  all <- c()

  while(step <= num_simulations){
    #print(step)

    # 2 pick a random node
    i <- sample(1:n,1)

    # 3 pick a neighbor
    neighbors <- which(network[i,] > 0)
    ps <- probas[i,neighbors]
    if(length(neighbors) > 1){
      j <- sample(neighbors, prob = ps, size=1)
    } else if(length(neighbors) == 1){
      j <- neighbors
    } else {
      next
    }

    # 6 update network of repetition
    lci <- lastcontacts[[i]]
    lci <- c(j,lci)
    lci <- lci[1:min(window_repetition,length(lci))]
    lastcontacts[[i]] <- lci

    lcj <- lastcontacts[[j]]
    lcj <- c(i,lcj)
    lcj <- lcj[1:min(window_repetition,length(lcj))]
    lastcontacts[[j]] <- lcj

    # 7 update probas
    ps <- probas_fixed[i,neighbors]
    t <- table(lci)
    newterms <- t[match(neighbors, names(t))]
    newterms[is.na(newterms)] <- 0
    ps <- ps * exp(param_repetition*newterms)
    probas[i,neighbors] <- as.vector(ps / sum(ps))

    neighbors <- which(network[j,] > 0)
    ps <- probas_fixed[j,neighbors]
    t <- table(lcj)
    newterms <- t[match(neighbors, names(t))]
    newterms[is.na(newterms)] <- 0
    ps <- ps * exp(param_repetition*newterms)
    probas[j,neighbors] <- as.vector(ps / sum(ps))


    # Entropy
    cpt_burnin <- cpt_burnin + 1
    if(cpt_burnin > burnin) cpt_thining <- cpt_thining + 1
    if(cpt_burnin > burnin && cpt_thining == thining){
      fakeprobas <- probas
      fakeprobas[fakeprobas == 0] <- 1
      fakedegrees <- degrees
      fakedegrees[fakedegrees == 1] <- 2
      allRH <- 1 + rowSums(probas[!isolates,] * log2(fakeprobas[!isolates,])) /
log2(fakedegrees[!isolates])
      print(mean(allRH))
      averageRH <- averageRH +mean(allRH)

      ntotal <- ntotal + 1
      cpt_thining <- 0
      all <- c(all,mean(allRH))
    }

    step <- step + 1
  }

  #plot(all)

  return(list(entropies = all,
              lastcontacts = lastcontacts))
```

```
}


# Umbrella functions for the optimization
optim_entropy_T <- function(p1,network,triads,homophilies,entropy_base){
  newentropy <- calculate_entropy_fixed(network,
                                        triads,
                                        homophilies,
                                        param_triad = p1,
                                        param_homophily = 0)
  return(abs(newentropy - entropy_base))
}
optim_entropy_H <- function(p2,network,triads,homophilies,entropy_base){
  newentropy <- calculate_entropy_fixed(network,
                                        triads,
                                        homophilies,
                                        param_triad = 0,
                                        param_homophily = p2)
  return(abs(newentropy - entropy_base))
}
```

R-script “script_net1000.R”

```
#########################################################
### Simulation of social interactions with infections ###
### Supplementary to: "Efficient social distancing    ###
### strategies to reduce the spread of COVID"         ###
### Authors: Marion Hoffman, Per Block                ###
#########################################################

library(igraph)
library(ggplot2)
library(Matrix)
source("functions_SN_covid_spread.R")

# Steps of the script:
# 1- generate network
# 2- Pre-calculate statistics
# 3- Calculate the entropy of the model with repetition strategy
# 4- Estimate parameters for the models with homophily and triadic strategies
# 5- Simulate random models and the three strategies in parallel


# 1- generate network

net_sim <- socialCirclesNet(nActors = 1000,
                            groupSize = 8, densityInGroups = 0.9,
                            minContactsGeo = 4, multMaxContactsGeo = 3, densityGeo
= 0.3,
                            randomTiesPerPerson = 0.5,
                            minContactsHomo   =   4,   multMaxContactsHomo   =   3,
homoStrength = 2,
                            orderAndUnique = T)

edgelist <- net_sim[[1]]
graph <- graph_from_edgelist(edgelist, directed = F) # igra
network <- get.adjacency(graph)
attributes <- net_sim[[2]]


# 2- Pre-calculate statistics
triads <- calculate_triads(graph,network)
homophilies <- calculate_homophilies(network,attributes)


# 3- Calculate the entropy of the model with repetition strategy
params_R <- c(0,0,2.5,2) # params: 1: triad, 2: homophily, 3: repetition, 4: window
repetition
num_simulations <- 50000
thining <- 500
burnin <- 5000
entropies_base <- calculate_entropy_timedependent(network,
                                                  triads,
                                                  homophilies,
                                                  param_triad = 0,
                                                  param_homophily = 0,
                                                  param_repetition = 2.5,
                                                  window_repetition = 2,
                                                  num_simulations                 =
num_simulations,
                                                  thining = thining,
                                                  burnin = burnin,
                                                  seed = 1)
entropy_base <- mean(entropies_base$entropies)
lastcontacts_base <- entropies_base$lastcontacts


# 4- Estimate parameters for the models with homophily and triadic strategies
pT <- optim(1, optim_entropy_T,
            network=network,
            triads=triads,
            homophilies=homophilies,
            entropy_base=entropy_base,
            lower=0, upper=40, method="L-BFGS-B")
params_T <- c(pT$par,0,0,0)
pH <- optim(1, optim_entropy_H,
            network=network,
```

```
            triads=triads,
            homophilies=homophilies,
            entropy_base=entropy_base,
            lower=0, upper=40, method="L-BFGS-B")
params_H <- c(0,pH$par,0,0)


# 5- Simulate random models and the three strategies in parallel
percentage_keepcontact <- 0.5
window_recovery <- 5
starting_nodes <- c(1,2)
seed <- 1234
burnin <- 0

res_random  <-  simulate(network,  triads,  homophilies,  percentage_keepcontact,
c(0,0,0,0),
                         window_recovery, starting_nodes, burnin = NULL, lastcontacts
= NULL, seed)
res_T <- simulate(network, triads, homophilies, percentage_keepcontact, params_T,
                  window_recovery,  starting_nodes,  burnin  =  NULL,  lastcontacts  =
NULL, seed)
res_H <- simulate(network, triads, homophilies, percentage_keepcontact, params_H,
                  window_recovery,  starting_nodes,  burnin  =  NULL,  lastcontacts  =
NULL, seed)
res_R <- simulate(network, triads, homophilies, percentage_keepcontact, params_R,
                  window_recovery, starting_nodes, burnin = burnin, lastcontacts =
lastcontacts_base, seed)

# n infected
sum(res_random$infected == -2)
sum(res_R$infected == -2)

# infection curve
plot(x = 1:res_random$number_steps_max, y = res_random$infections, type = "l")
plot(x = 1:res_R$number_steps_max, y = unlist(res_R$infections), type = "l")
```