Combining fine-scale social contact data with epidemic modelling reveals interactions between contact tracing, quarantine, testing and physical distancing for controlling COVID-19

Josh A. Firth[1,2], Joel Hellewell[3], Petra Klepac[3,4], Stephen Kissler[5], CMMID COVID-19 working group, Adam Kucharski[3], Lewis G. Spurgin[6*]

1. Department of Zoology, University of Oxford, Oxford, UK
2. Merton College, University of Oxford, Oxford, UK
3. Centre for the Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK
4. Department for Applied Mathematics and Theoretical Physics, University of Cambridge
5. Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston MA
6. School of Biological Sciences, University of East Anglia, Norwich, UK

***Correspondence:** l.spurgin@uea.ac.uk

CMMID COVID-19 working group members (order selected at random): Mark Jit, Katherine E. Atkins, Samuel Clifford, C Julian Villabona-Arenas, Sophie R Meakin, Charlie Diamond, Nikos I Bosse, James D Munday, Kiesha Prem, Anna M Foss, Emily S Nightingale, Kevin van Zandvoort, Nicholas G. Davies, Hamish P Gibbs, Graham Medley, Amy Gimma, Stefan Flasche, David Simons, Megan Auzenbergs, Timothy W Russell, Billy J Quilty, Eleanor M Rees, Quentin J Leclerc, W John Edmunds, Sebastian Funk, Rein M G J Houben, Gwenan M Knight, Sam Abbott, Fiona Yueqian Sun, Rachel Lowe, Damien C Tully, Simon R Procter, Christopher I Jarvis, Akira Endo,

Kathleen O'Reilly, Jon C Emery, Thibaut Jombart, Alicia Rosello, Arminder K Deol, Matthew Quaife, Stéphane Hué, Yang Liu, Rosalind M Eggo, Carl A B Pearson

Abstract

Case isolation and contact tracing can contribute to the control of COVID-19 outbreaks. However, it remains unclear how real-world networks could influence the effectiveness and efficiency of such approaches. To address this issue, we simulated control strategies for SARS-CoV-2 in a real-world social network generated from high resolution GPS data. We found that tracing contacts-of-contacts reduced the size of simulated outbreaks more than tracing of only contacts, but resulted in almost one third of the local population being quarantined at a single point in time. Testing and releasing non-infectious individuals reduced the numbers of quarantined individuals without large increases in outbreak size, but high testing rates were required for this to be effective. Finally, if testing availability is constrained, we estimated that combining physical distancing with contact tracing could enable epidemic control while reducing the number of quarantined individuals. Our approach highlights the importance of network structure and social dynamics in evaluating the potential impact of SARS-CoV-2 control.

Introduction

Non-pharmaceutical interventions (NPIs) remain central to reducing SARS-CoV-2 transmission [1–3]. Such responses generally include: case isolation, tracing and quarantining of contacts, use of PPE and hygiene measures, and policies designed to encourage physical distancing (including closures of schools and workplaces, banning of large public events and restrictions on travel). Due to the varying economic and social costs of these non-pharmaceutical interventions, there is a clear need for sustainable strategies that limit SARS-CoV-2 transmission while reducing disruption as far as possible.

Isolation of symptomatic cases, and quarantine of their contact (e.g. household members), is a common public health strategy for reducing disease spread[4,5]. This approach has been used as part of SARS-CoV-2 control strategies[6]. However, the relatively high reproduction number of the SARS-CoV2 virus in early outbreak stages[7,8], alongside likely high contribution to transmission from presymptomatic and asymptomatic individuals[9], means that manual tracing of contacts alone may not be a sufficient containment strategy under a range of outbreak scenarios[10]. As countries relax lockdowns and other more stringent physical distancing measures, combining the isolation of symptomatic individuals and quarantine of contacts identified through fine-scale tracing is likely to play a major role in many national strategies for targeted SARS-CoV-2 control[11]. Modelling studies suggest that app-based tracing can be highly effective as a containment strategy if uptake is high (~80% of smartphone users) and that very large numbers of individuals could potentially be quarantined. However, these results, along with those more generally regarding COVID-19 transmission, rely primarily on simulating or assuming the structure of fine-scale social networks, and this may not accurately capture the effect of contact-based interventions[12]. To fully understand how contact tracing may be effectively combined with other physical distancing measures to enable containment, while reducing the number of people in quarantine, therefore requires realistic data on social network structure.

It is possible to assess the potential effectiveness of contact tracing by simultaneously modelling disease spread and contact tracing strategies through social systems of individuals[13]. These systems are usually simulated through parameterisation with simple social behaviours (e.g. the distribution of the number of physical contacts per individual). Further still, social systems may be simulated as networks that can be parameterised according to assumptions regarding different contexts (for example, with different simulated networks for households, schools and workplaces), or using estimated contact rates of different age groups[14]. However, much less is known about how different types of real-world social behaviour and the hidden structure found in real-life networks may affect both patterns of disease transmission and efficacy of contact tracing under different scenarios[15,16]. Examining contagion dynamics and control strategies using a ‘real-world’ network allows for a more realistic simulation of SARS-CoV-2 outbreak and contact tracing dynamics.

Datasets recording detailed social interactions amongst people are rare, and social networks are instead commonly inferred from either self-reported contacts (which rely on recall accuracy and may miss some contact events) or limited tracking data within single settings such as schools and workplaces (and therefore missing contact events in other contexts and ignoring bridging between contexts). One of the most comprehensive accessible datasets on human social interactions collected specifically for modelling infectious disease dynamics was generated by a 2017/18 citizen science project as part of the British Broadcasting Corporation (BBC) documentary “Contagion! The BBC Four Pandemic”. The high-resolution data collection focused on residents of the town of Haslemere, where the first evidence of UK-acquired infection with SARS-CoV-2 would later be reported in late February 2020[17]. Previous analyses of this dataset have shown that it is structurally relevant to modelling disease spread, and hence holds substantial potential for understanding and controlling real-world diseases[18,19]. Combining this dataset with infectious disease transmission modelling offers a unique opportunity to understand how NPIs can be best implemented to contain SARS-CoV-2.

Here we develop an epidemic model which simulates COVID-19 outbreaks across the Haslemere network, and assess the impact of a range of testing and contact tracing strategies for controlling these outbreaks. We then simulate 'test and release' strategies and physical distancing strategies and quantify how the interaction between physical distancing, contact tracing and testing affects outbreak dynamics.

Methods

Social tracking data

The Haslemere dataset was generated and described as part of previous work[18,19]. Briefly, the data were collected during the 2017/18 *BBC Pandemic* project conducted in Haslemere, Surrey, UK. The project involved a massive citizen-science experiment to collect social contact and movement data using a custom-made phone app, and was designed to generate data relevant to understanding directly transmitted infectious disease[18,19]. Of the 1272 individuals within Haslemere that downloaded the app, 468 individuals had sufficient data points at a resolution of 1m over three full days within the focal area for further analysis[18]. The dataset used here includes these 468 individuals, with de-identified proximity data made available as pairwise distances (~1 m resolution) at 5 min intervals (excluding 11pm-7am)[18].

Social network construction

In our primary analysis, we defined social contacts as events when the pairwise distances between individuals within a 5 min time interval (calculated using the Haversine formula for great-circle geographic distance[18]) are 4 m or less. By doing so, we aimed to capture the majority of relevant face-to-face contacts (i.e. those that may result in transmission) over 5 min periods, particularly given the 1 m potential error[18] on the tracking measurement during these short time intervals. Furthermore, this 4 m threshold is within typical mobile phone Bluetooth ranges for relatively

accurate and reliable detections. Therefore, this contact dataset will also be comparable to proximity-based contacts identified through Bluetooth contact tracing apps, which may be preferred to real-location tracking for privacy reasons. We considered the sensitivity of the network to the contact definition by testing six further social networks from contacts defined using different threshold distances spanning the conceivable potential transmission range within the 5 min intervals (1 m to 7 m thresholds). We first measured the correlation of the network structure (i.e. pairwise contacts) across the seven networks using Mantel tests. We also measured the correlation of each individual's degree (number of contacts), clustering coefficient (number of contacts also connected to one another), betweenness (number of shortest paths between nodes that pass through an individual), and eigenvector centrality (a measure that accounts both for a node's centrality and that of its neighbours) across the seven networks.

The Haslemere data is a temporal dataset spanning three full days. While the epidemic model we use is dynamic (see below Methods), the contagion process of COVID-19 operates over a longer time period than three days. To be able to meaningfully simulate longer-term outbreak dynamics, we quantified the data as a static social network in which edges indicate the propensities for social contact between nodes. Temporal information is incorporated by weighting the edges using the temporal contact information, instead of using a dynamic network within the dynamic model, as this would require contact data over a much longer period. In the primary analysis, we weighted the edges as the number of unique days a dyad was observed together (but see Supplementary Information for other temporal definitions). Therefore, the weight score indicates the propensity for each dyad to engage in a social contact event on any given day, whereby 0 = no contact, 1 = ‘weak links’ observed on the minority of days (one third), 2 = ‘moderate links’ observed on the majority of days (two thirds), and 3 = ‘strong links’ observed on all days. In this way, the weights of this social network could be included directly, and intuitively, into the dynamic epidemic model (see below). For sensitivity analysis, we also created networks and examined the correlation in dyadic social associations scores (using Mantel tests). We used edges specified as i) a binary (i.e. unweighted)

network across all days, ii) a raw (and ranked) count of 5 min intervals in contact, iii) a transformed weighted count (edge weight transformed as $1 - e^{interval\ count}$, which approximates a scenario where infection risk increases until ~30mins of contact between dyads) and iv) a 'simple ratio index' (SRI) weighting that corrects for observation number as SRI score[20]. The SRI score for any two individuals (i.e. A and B) is calculated as:

(1) $$SRI_{A,B} = \frac{Obs_{A,B}}{Obs_A + Obs_B - Obs_{A,B}} ,$$

where *Obs* is the number of 5 min observation periods (the intervals since the start of the day) within which an individual is recorded within 4 m of another individual.

Null network simulation approach

We used null networks[21] to understand the network properties that shape predictions of COVID-19 spread under different control scenarios. Null networks can also show how contagion may operate in different social environments, and which simulation approaches may be the most similar to real-world infection dynamics. We created four null network scenarios (Fig. S1) with 1000 networks generated under each of these. All of the null network scenarios kept the same number of nodes, edges, and weights of these edges, as the Haslemere network, but were generated under the following nulls: (1) 'edge null' (Fig. S1A) considered random social associates, allowing the edges of the network to be randomly allocated between all nodes; (2) 'degree null' (Fig. S1B) considered individual differences in sociality but random social links between dyads, so randomly swapped the edges between nodes but maintained the degree distribution of the real network (and was, therefore, even more conservative than a power-law network simulation aiming to match real differences in sociality); (3) 'lattice null' (Fig. S1C) considered triadic and tight clique associations, so created a ring-like lattice structure through assigning all edges into a ring, with individuals connected to their direct neighbours, and those of the second and third order (i.e. six links per

individual) and then randomly removing excess links; (4) ‘cluster null’ (Fig. S1D) considered the observed level of clustering, so created a ring structure as described above but only between individuals observed as connected (at least 1 social link) in the real network, added remaining links (sampled from 4th order neighbours), and then rewired the edges until the real-world global clustering was observed (~20% rewiring; Fig. S1D). These conservative (and informed) null models allowed connections to be arranged differently within the network but maintained the exact same number of individuals, social connections and weights of these social connections at each simulation.

Epidemic model

Building on the epidemiological structure of a previous branching-process model[10], we developed a full epidemic model to simulate COVID-19 dynamics across the Haslemere network. Full model parameters are given in Table 1. For a given network of individuals, an outbreak is seeded by randomly infecting a given number of individuals. All newly infected individuals are assigned an ‘onset time’ drawn from a Weibull distribution that determines the point of symptom onset (for symptomatic individuals), and the point at which infectiousness is highest (for all individuals). Each individual is then simultaneously assigned asymptomatic status (whether they will develop symptoms at their onset time), as well as presymptomatic status (whether or not they will infect before their assigned onset time), drawn from Bernoulli distributions with defined probabilities (Table 1). At the start of each day, individuals are assigned a status of susceptible, infectious or recovered (which would include deaths) based on their exposure time, onset time and recovery time (calculated as onset time plus seven days), and are isolated or quarantined based on their isolation/quarantine time (described below). The model then simulates infection dynamics over 70 days.

Possible infectors are all non-isolated and non-quarantined infectious individuals. Each day, all susceptible contacts of all infectors within the network are at risk of being infected. The transmission rate for a given pair of contacts is modeled as:

(2) $$\lambda(t, s_i, p_i) = A_{s_i} I_{ei} \int_{t-1}^{t} f(u;\ \mu_i, \alpha_{p_i}, \omega_{p_i}) du$$

where t is the number of days since the infector *i* was exposed, s_i and p_i are the infector's symptom status (asymptomatic yes/no, and presymptomatic yes/no, respectively). A_{s_i} is the scaling factor for the infector's symptomatic status (Table 1) and I_{ei} is the weighting of the edge in the network (i.e. number of days observed together) between the infector and the susceptible individual. The probability density function $f(u;\ \mu_i, \alpha_{p_i}, \omega_{p_i})$ corresponds to the generation time, which is drawn from a skewed normal distribution (see [10] for details). Briefly, this uses the infector's onset time as the location parameter μ_i, while the slant parameter α_{p_i} and the scale parameter ω_{p_i} both vary according to the infector's presymptomatic transmission status (Table 1). This enabled us to simulate a predefined rate of presymptomatic transmission, while retaining a correlation structure between onset time and infectiousness, and avoiding a scenario whereby a large number of individuals were highly infectious on the first day of exposure (see Table 1 and data sharing for more details).

Using this transmission rate, the probability of infection between a susceptible-infected pair of individuals *t* days after the infector's exposure time is then defined as:

(3) $$P(t, s_i, p_i) = 1 - e^{-\lambda(t, s_i, p_i)}$$

Note that the recovery time threshold of seven days does not affect infection dynamics (as transmission rate ≈ 0 seven days after onset time), but is instead used for contact tracing purposes (see below). To test how the above rate of infection related to the reproduction number R_0 and the observed generation times, we generated empirical estimates of the number of secondary infections in the early outbreak stages of the model. We ran 1000 trial simulations from a random single starting infector and quantified i) the mean number of secondary infections from this case, and ii) the time at which each secondary case was infected. We found that the above equation corresponded to R_0 = 2.8, and a mean generation time of 5.6 days (median = 5 days), which correspond closely to recent estimates[9,22]. Nonetheless, we performed sensitivity analysis of R_0 by multiplying the rate of infection by a scaling parameter (Table 1).

In addition to the infection rate from within the network, the infection rate from outside the network is also simulated daily by randomly infecting susceptible individuals with a probability of 0.001 (although we also performed sensitivity analysis of this parameter).

We simulated different contact tracing scenarios using contact information from the network, with the aim of evaluating both app-based and manual contact tracing strategies. Primary and secondary contacts of individuals are identified from the network on the day of the infector's symptom onset and, as such, contacts of asymptomatic infectors are not traced. Contacts who have already recovered are excluded. Susceptible contacts are traced with a given probability (0.4-0.8 tested - see table 1). We assume that this probability captures a wide range of reasons why contacts might not be traced, and it thus acts as an intuitive simplification.

The isolation and/or quarantine time of each individual is determined based on their infection status, their symptomatic status, whether they have been traced, and the control scenario. We consider four control scenarios: i) no control, where no individuals are isolated or quarantined, ii) case isolation, where individuals isolate upon symptom onset after a delay period, iii) primary

contact tracing with quarantine, where individuals isolate upon symptom onset (after a delay) and traced contacts are quarantined upon their infector's symptom onset (also after a delay), and iv) secondary contact tracing, as scenario iii) but including contacts of contacts. All isolated and quarantined individuals are contained for 14 days.

Finally, we simulated a range of testing efforts for SARS-CoV-2. Each individual is assigned a testing time on isolation or quarantine, with the delay between containment and testing sampled from a Weibull distribution. A cap of the maximum number of daily tests is assigned, and each day up to this number of individuals are randomly selected for testing. Test results are dependent on infection and asymptomatic status, with a false negative rate (i.e. the probability that an infectious case will test negative) of 0.5 for asymptomatic patients and 0.1 for symptomatic cases[23], and a false positive rate (i.e. the probability that susceptible case will test positive) of 0.02[24]. Cases who tested negative were immediately released from isolation/quarantine.

A set of default parameters were chosen to represent a relatively optimistic model of contact tracing, which included a short time delay between symptom onset/tracing and isolation/quarantine (1-2 days), and a high proportion (80%) of contacts traced within this tracked population (default parameters highlighted in bold in Table 1). We assumed that the probability of tracing was constant over time, and therefore independent of previous isolation/quarantine events, and that all individuals remained in quarantine for the full 14 days, unless released via testing. We performed sensitivity tests on all relevant parameters (Table 1).

We ran each simulation for 70 days, at which point the majority of new infections came from outside the network (see results), with all scenarios replicated 1000 times. With the null networks (above) and physical distancing simulations (below), we ran one replicate simulation on each of 1000 simulated networks. Due to the finite population size and nature of the Haslemere dataset, in no simulations were all individuals in the population infected under our default settings. Therefore,

for each simulation we report the number of cases per week, and quantify the total number of cases after 70 days as a measure of outbreak severity. To present the level of isolation and quarantine required under different scenarios, we calculate the number of people contained on each day of the outbreak, and average this over the total number of weeks to get weekly changes in the rates of isolations and quarantines per day.

Physical distancing Simulations

We simulated a population-level physical distancing effort, whereby a given proportion of the 'weak links' (edges only observed on a single day) were removed but then randomly reassigned to remaining 'weak links' or 'moderate links' (edges observed on two days). The benefit of this simulation technique is that the overall weighted connectivity of the network remains unchanged (equal to edges*weights) but the number of unique edges is decreased to the number of weak links selected to be removed (Fig S2A-D). This is somewhat akin to a simple situation whereby individuals reduce their unique contacts (e.g. to people outside of their household) and instead engage in more social contacts with their remaining associates (e.g. those inside their household). We also carried out a more complex physical distancing simulation, whereby the probability of a weak link being selected for removal was inversely proportional to the raw count of 5 min interval connections observed for that dyad, and the reassignment of these links to remaining edges was proportional to the amount of time dyads were together (Fig S2E-G). This represents a scenario where individuals stop contact with people they spend the least time with, and reallocate this social time to the contacts that they already spend the most time with.

The epidemic model code can be accessed at: https://github.com/biouea/covidhm

Results

Social network properties and dynamics

Defining dyadic contacts on a day-by-day basis as at least one daily 5 min period with a distance of 4 m (see Methods) gave 1616 daily contact events and 1257 unique social links between individuals. The social network defined in this way was significantly and strongly correlated (r >0.85 in all cases) with social networks based on other contact distances (from 1-7 m contact ranges) for defining contacts (Fig. S3). Similarly, social networks created using different time-periods for weighting the dyadic contacts (Fig. S4) were also strongly related to the weighting used here (i.e. number of days seen together). As such, this social network quantification not only gives a representative indication of daily contact propensities within the relevant transmission range between individuals (see Methods) but also captures much of the patterns and structure presented by different quantifications of this social system.

Epidemic model and control scenarios

Example outbreaks across the Haslemere network of 468 individuals under different control scenarios are displayed in Fig. 1, with a full animated visualisation in Supplementary Video 1, and a Shiny app is available to run individual outbreak simulations (see data sharing). Starting with a single infected individual (Fig. 1A), scenarios with no control measures quickly led to substantial numbers of infections (Fig. 1B-D), while contact tracing scenarios reduced the number of infections but resulted in a large number of contained cases in early-mid outbreak stages (Fig. 1E-G). Across all simulations, our epidemic model showed that uncontrolled outbreaks in the Haslemere network stemming from a single infected individual resulted in a median of 12% (IQR = 9.4%-15.8%) of the population infected after 70 days (Fig. 2). Isolation when symptomatic resulted in 9.3% (7.9%-11.3%) of the population infected, while primary contact tracing resulted in 9% (7.7%-10.5%) of infected. Secondary contact tracing resulted in the largest reduction (7.3%, 6.4%-8.3%) of the population infected after 70 days. The number of quarantined individuals was very high under both primary and secondary contact tracing, with a median of 29% (IQR = 19%-40%) of the population quarantined during the outbreak peak with the latter (Fig. 2). Examining temporal dynamics showed that outbreak peaks typically occurred within the first 1-3 weeks, and that interventions

reduced the overall size of the outbreaks as well as their growth rate (Fig. 2). The number of people required to isolate or quarantine followed a similar trajectory to the number of cases, although under secondary contact tracing, substantial proportions of the population (13%, 7%-20%) were quarantined even at the end of the simulations (Fig. 2).

Null network models based on the Haslemere data (and all of which maintained the exact number of individuals, connections and weights of connections, but shuffled network architecture), largely confirmed overall patterns found with the real-world network, albeit with some important differences (Fig. 3). The number of cases estimated using the null networks was broadly similar to the real-world network, although the ‘lattice’ and ‘cluster’ networks both slightly underestimated the number of cases relative to the real-world network. Importantly, the rate of quarantine varied substantially among the null networks, especially under secondary contact tracing (Fig. 3). In particular, the clustered and lattice networks both substantially underestimated the number of quarantined cases, while the ‘degree null’ network overestimated the number of quarantined cases (Fig. 3). Together, this demonstrates the importance of social network structure in shaping this contagion and the effectiveness of control measures.

Sensitivity analysis of the efficacy of contact tracing under the epidemic model is presented in Figures S5-S10. As expected, outbreak size decreased with the percentage of contacts traced in all scenarios, and increased with the reproduction number (Fig. S5), the proportion of asymptomatic cases (Fig. S6), the proportion of pre-onset transmission (Fig. S7), the delay between onset/tracing and isolation/quarantine (Fig. S8), and the number of initial cases (Fig. S9). These parameters also had an effect on the number of isolated contacts. For instance, a reduced delay time between onset and case/contact isolation resulted in not only a reduction in the number of cases, but also a reduction in the number of contacts required to be traced (Fig. S8). Simulating a range of outside infection rates showed that this parameter had a large effect on the number of cases, which increased with outside infection rate across all intervention scenarios, as did the

number of isolated cases (Fig. S10). The tradeoff between the number of cases and the number of quarantined cases was found across the entirety of the parameter space (Figs S5-S10).

We also assessed how the testing and releasing of isolated and quarantined subjects might affect the numbers of cases and time spent in isolation and quarantine, while considering false positive and negative rates. We estimated that increasing the testing capacity (and therefore testing and releasing more quarantined cases) led to only very small increases in the outbreak size (median 7.9%, IQR 6.8%-9.6%; Fig. 4). However, high levels of testing led to a substantial reduction in the number of quarantined cases in both primary and secondary contact tracing scenarios, with on average 1.7% (0.7%-3.3%) and 11.7% (6%-22%) quarantined cases during the outbreak peaks, respectively, when testing capacity was 50 tests per day. However, the number of tests required to reduce the numbers of quarantined cases were large, especially under secondary contact tracing, where 19% of the population (IQR 6%-22%) required tests in a single week during outbreak peaks (Fig. 4).

We simulated physical distancing by reducing the number of weak links in the Haslemere network, while retaining the same overall number of social interactions. We found that, across control scenarios, physical distancing led to only a small reduction in the number of overall cases (Fig. 5). Importantly however, increasing physical distancing was associated with marked reductions in the number of quarantined cases under both primary (1.2%, 0.5%-2.2%) and secondary contact tracing (5.2%, 2.7%-8.7%), as well as reducing the number of tests required (Fig. 5). Simulating physical distancing using an alternative approach based on the amount of time spent with contacts within days (see methods) yielded qualitatively identical results to our simpler model (Fig. S11).

Discussion

Through assessing the predicted spread of COVID-19, alongside interactions between disease control methods, this study highlights a number of challenges, and some promising ways forward,

for medium-term control of SARS-CoV-2. Although several studies have attempted to predict the efficacy of interventions, and contact tracing in particular, for controlling the SARS-CoV-2 pandemic[10–12], we have limited understanding of how contact tracing might affect SARS-CoV-2 transmission dynamics in the real world[10–12]. Compared to previous models, the present study examines transmission dynamics in a real-world network, but over a relatively small geographical area. Further, the number of infections found in our epidemic models were reasonably low, which might reflect the fact that the Haselmere dataset represents a sample of a larger population. As such, while our study offers new insight into local populations, we do not know to what extent the dynamics found here will extrapolate larger-scale social systems. Nonetheless, we have demonstrated that important trade-offs exist when intervention methods are applied within local populations.

In regards to the effectiveness of strategies, our model corroborates with models using simulated social systems and showing that, for a disease such as COVID-19 with high levels of transmission from asymptomatic and presymptomatic individuals, contact tracing is likely to be most effective when the proportion of traced contacts is high, when the delay from notification to quarantine is short[10], and, most importantly, when the number of starting cases and rate of movement into the network are low. In all scenarios, the tracing and quarantining of contacts resulted in fewer cases than case isolation alone, with the tracing of secondary contacts leading to fewer cases than primary tracing. Importantly, however, regardless of model parameters, contact tracing is only effective because it results in a very large number of people being quarantined (Fig. 2). This is in line with a large-scale recent simulation model of app-based contact tracing in the UK[12]. Further, in our (optimistic) default parameter settings we assumed that 20% of contact tracing attempts were missed. This, combined with the very large number of quarantined cases under secondary contact tracing (Fig. 2), suggests that a majority of the population could receive a notification that they should quarantine within the first 2-3 weeks of an outbreak.

The number of quarantined cases can be reduced through mass testing and release of individuals who return a negative result. Conversely, due to the high false negative rates associated with RT-PCR tests for SARS-CoV-2[23,25], large-scale test and release strategies could result in missing positive cases and decrease the effectiveness of contact tracing. In our model, increasing the rate of testing (and release) of cases led to a reduction in the number of people quarantined with only a small increase in final outbreak size (Fig. 4), despite incorporating the relatively high false-negative rates observed, especially in asymptomatic cases[23]. However, we assumed a short delay between isolation/quarantine and testing, so our results on testing only apply to situations where testing of quarantined cases can be carried out rapidly (in less than 2 days). Further, the secondary contact tracing scenarios which resulted in the largest reduction in outbreak size were associated with a very large number of quarantined contacts (Fig. 2). Accordingly, a very high testing rate would be required to reduce the number of quarantined cases (Fig. 4), with up to a fifth of the population requiring tests in a single week during outbreak peaks. Again, we cannot be sure to what extent our results will represent larger populations, but the tripartite relationship between the number of cases, the number of quarantined contacts and the number of tests required will apply in the majority of scenarios in which rates of social interaction are high.

A very high notification and quarantine rate for any contact tracing system may have consequences for adherence. Our model is optimistic in its assumption that individuals isolate independently of previous notifications or isolations, and highly optimistic in its assumption of 100% adherence to quarantine among traced contacts. In reality a high notification and quarantine rate may result in individuals being less likely to undertake quarantine in the future, which in turn will impact outbreak dynamics. It has been suggested that this can be addressed through (digital) targeted quarantine requests to the individuals at highest risk of infection, or to those most likely to spread to others[26], which could be addressed in future studies using the framework and methodology presented here. The likely effectiveness of these approaches in terms of reducing outbreak size and keeping quarantine rates low is an important area of ongoing research.

In the absence of (or in addition to) targeted contact tracing approaches, combining contact tracing with other physical distancing measures may allow for outbreak control while reducing the number of people in quarantine, and the number of tests required (Fig. 5). We aimed to consider low to moderate levels of physical distancing, so used a model whereby the number of social interactions remains the same, but interactions with ‘rare’ contacts are reassigned to ‘common’ contacts. We do not have information on household structure within the Haslemere dataset, but our physical distancing scenario is analogous to decreasing the level of non-household contacts and increasing the level of household contacts. The number of cases, as well as the number of quarantined individuals and tests required, decreased only slightly with the degree of physical distancing. Importantly, when physical distancing was in place, the difference in the number of cases between primary and secondary contact tracing strategies was small, yet primary contact tracing resulted in fewer quarantined cases and fewer tests required (Fig. 5). It may therefore be the case that when physical distancing measures are in place and contact rates are relatively low, primary contact tracing is the most efficient strategy. Further work is required to determine exactly what kinds of physical distancing measures would enable effective outbreak control alongside contact tracing.

Network structure can have substantial effects on epidemic model predictions[27,28], and our null network modelling approach shows that this is important when considering SARS-CoV-2 spread and the effectiveness of control measures within real-world structures. Indeed, each null model maintained the same number of people as the real network, and the same number and strength of social associations, but simply reordered the connections between individuals. This reordering alone changed in the emerging predictions, thus highlighting the importance of the fine-scale arrangement of social connections. Specifically, null models that randomised the order of social connections, and those that incorporated information on the distribution of ties between individuals, appeared to match the real network predictions best. On the other hand, the lattice and clustered networks both underestimated outbreak size, and substantially overestimated the effectiveness of

contact tracing. These results demonstrate that the use of network-based simulations of SARS-CoV-2 dynamics requires caution, as even if such models had precise information on the number of individuals and amount of social interactions occurring within a system, the assumed architecture of the social network structure alone can shape predictions for both the extent of spread and the usefulness of control strategies. Furthermore, through providing insight into how changes to network structure influences contagion dynamics, the null network simulation approach gives some indication of how this contagion and associated control strategies may operate in different social environments. For instance, different social structures may arise when considering particular social settings (e.g. workplaces, commuting), some of which may be closer to the extreme random edge null networks generated here, while others may represent the lattice or clustered null networks. Considering this structure will lead to improved predictions of outbreak dynamics.

There are a number of important limitations to our study and the current availability of empirical data in general. Most importantly, this social network is taken from a single, small town and over a short period of time and we do not know to what extent the social dynamics will be applicable to larger cities and other contexts and over long periods. Therefore, future large-scale efforts in gathering data on dynamic fine-scale social behaviour over long periods of time (ideally over the entire contagion period) in major cities would be of great benefit for assessing the relative uses of SARS-CoV-2 control strategies. Further, the Haslemere data, while rich, does not sample the entire population of Haslemere, and children under the age of 13 were not included in the experiment, which could potentially have an impact on outbreak and social tracking dynamics. Again, such issues are also likely to be prevalent across real-world contact-tracing attempts, as the ability to track children will be limited, particularly with app-based approaches that require a smartphone. It is encouraging that our results broadly align with other, larger-scale simulations of contact tracing which explicitly model these limitations, but lack the fine-scale social tracking data[12]. Therefore, by supplying a general framework for simulating the spread of COVID-19 on

real-world networks, we hope to promote integration of multiple real-world social tracking datasets with epidemic modelling, which may provide a promising way forward for optimising contact tracing strategies and other non-pharmaceutical interventions.

Acknowledgements

This work was instigated through the Royal Society's Rapid Assistance in Modelling the Pandemic (RAMP) scheme. We thank Michael Pointer for helpful discussions throughout, and Cock van Oosterhout and Julia Gog for comments on the manuscript. We thank all those in Haslemere who took part in the BBC Pandemic study. We thank Hannah Fry and 360 Production, especially Danielle Peck and Cressida Kinnear, for making possible the collection of the dataset that underlies this work, and Andrew Conlan, Maria Tang, and Julia Gog for their contribution to the BBC study. JAF was supported by a research fellowship from Merton College and BBSRC (BB/S009752/1) and acknowledges funding from NERC (NE/S010335/1). PK was in part funded by the Royal Society under award RP\EA\180004 and European Commission: 101003688. AJK was supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant Number 206250/Z/17/Z).

Data sharing

This study used the raw data previously published in Kissler et al. 2018 (made available with full description here: https://www.biorxiv.org/content/10.1101/479154v1)

The code and data used to produce the simulations is available as an R package at: https://github.com/biouea/covidhm

A shiny app which runs individual outbreak simulations is available at: https://biouea.shinyapps.io/covidhm_shiny/

References

1. Ferguson, N. *et al.* Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. (2020).
2. Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
3. Tian, H. *et al.* An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* (2020) doi:10.1126/science.abb6105.
4. Fraser, C., Riley, S., Anderson, R. M. & Ferguson, N. M. Factors that make an infectious disease outbreak controllable. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6146–6151 (2004).
5. Peak, C. M., Childs, L. M., Grad, Y. H. & Buckee, C. O. Comparing nonpharmaceutical interventions for containing emerging epidemics. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4023–4028 (2017).
6. Chen, S. What's behind Vietnam's coronavirus containment success? *South China Morning Post* https://www.scmp.com/news/asia/southeast-asia/article/3079598/coronavirus-whats-behind-vietnams-containment-success (2020).
7. Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* (2020) doi:10.1016/S1473-3099(20)30144-4.
8. Klinkenberg, D., Fraser, C. & Heesterbeek, H. The effectiveness of contact tracing in emerging epidemics. *PLoS One* **1**, e12 (2006).
9. He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* (2020) doi:10.1038/s41591-020-0869-5.
10. Hellewell, J. *et al.* Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob Health* **8**, e488–e496 (2020).

11. Ferretti, L. *et al.* Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* (2020) doi:10.1126/science.abb6936.
12. Hinch, R. *et al. Effective Configurations of a Digital Contact Tracing App: A report to NHSX*. https://github.com/BDI-pathogens/covid-19_instant_tracing (2020).
13. Eames, K. T. D. & Keeling, M. J. Contact tracing and disease control. *Proc. Biol. Sci.* **270**, 2565–2571 (2003).
14. Del Valle, S. Y., Hyman, J. M., Hethcote, H. W. & Eubank, S. G. Mixing patterns between age groups in social networks. *Soc. Networks* **29**, 539–554 (2007).
15. Kiss, I. Z., Green, D. M. & Kao, R. R. Disease contact tracing in random and clustered networks. *Proc. Biol. Sci.* **272**, 1407–1414 (2005).
16. Read, J. M., Eames, K. T. D. & Edmunds, W. J. Dynamic social networks and the implications for the spread of infectious disease. *J. R. Soc. Interface* **5**, 1001–1007 (2008).
17. BBC News. Coronavirus patient first to be infected in UK. *BBC* (2020).
18. Kissler, S. M., Klepac, P., Tang, M., Conlan, A. J. K. & Gog, J. R. Sparking 'The BBC Four Pandemic': Leveraging citizen science and mobile phones to model the spread of disease. *bioRxiv* 479154 (2018) doi:10.1101/479154.
19. Klepac, P., Kissler, S. & Gog, J. Contagion! The BBC Four Pandemic--The model behind the documentary. *Epidemics* **24**, 49–59 (2018).
20. Cairns, S. J. & Schwager, S. J. A comparison of association indices. *Anim. Behav.* **35**, 1454–1469 (1987).
21. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).
22. Davies, N. G. *et al.* The effect of non-pharmaceutical interventions on COVID-19 cases, deaths and demand for hospital services in the UK: a modelling study. *Infectious Diseases (except HIV/AIDS)* (2020) doi:10.1101/2020.04.01.20049908.
23. Chau, N. V. V. *et al.* The natural history and transmission potential of asymptomatic SARS-CoV-2 infection. *Infectious Diseases (except HIV/AIDS)* (2020)

doi:10.1101/2020.04.27.20082347.

24. Cohen, A. N. & Kessel, B. False positives in reverse transcription PCR testing for SARS-CoV-2. *Epidemiology* (2020) doi:10.1101/2020.04.26.20080911.
25. Li, Y. *et al.* Stability issues of RT-PCR testing of SARS-CoV-2 for hospitalized patients clinically diagnosed with COVID-19. *J. Med. Virol.* (2020) doi:10.1002/jmv.25786.
26. McCall, B. Shut down and reboot-preparing to minimise infection in a post-COVID-19 era. *Lancet Digit Health* (2020) doi:10.1016/S2589-7500(20)30103-5.
27. Keeling, M. J. & Eames, K. T. D. Networks and epidemic models. *J. R. Soc. Interface* **2**, 295–307 (2005).
28. Xu, Z. & Sui, D. Z. Effect of Small-World Networks on Epidemic Propagation and Intervention. *Geogr. Anal.* **41**, 263–282 (2009).
29. Backer, J. A., Klinkenberg, D. & Wallinga, J. The incubation period of 2019-nCoV infections among travellers from Wuhan, China. *medRxiv* 2020.01.27.20018986 (2020).
30. Donnelly, C. A. *et al.* Epidemiological determinants of spread of causal agent of severe acute respiratory syndrome in Hong Kong. *Lancet* **361**, 1761–1766 (2003).
31. Kliger, A. S. & Silberzweig, J. Mitigating Risk of COVID-19 in Dialysis Facilities. *Clin. J. Am. Soc. Nephrol.* **15**, 707–709 (2020).
32. Liu, Y., Centre for Mathematical Modelling of Infectious Diseases nCoV Working Group, Funk, S. & Flasche, S. The contribution of pre-symptomatic infection to the transmission dynamics of COVID-2019. *Wellcome Open Res* **5**, 58 (2020).

Tables and Figures

Table 1. Parameter values for the epidemic model. Numbers given for sampled parameters are medians and interquartile ranges, and default parameter settings for the scenario models are highlighted in bold.

Parameter	Assumed value(s)	Details and references
Sampled		
Incubation period	5.8 days (2.6)	[9,29]
Serial interval	Location = incubation period For post-symptomatic transmission, slant = ∞, scale = 2 For presymptomatic transmission, slant = -∞, scale = incubation period.	Based on data in [9]
Delay from onset/tracing to isolation, and from isolation to testing	**1 day (0.4-1.9) days ('short')** 3.5 days (2.8-5.2) days ('medium')	Assumed (short) and [30] (medium)
Fixed		
Initial cases	**1**, 5	Assumed
Scaling parameter (and corresponding empirical estimate of the reproduction number R_0)	**1 (2.8)**, 2 (3.5)	[31]
Percentage asymptomatic individuals	20%, **40%**	[12]
Infectiousness of asymptomatic individuals	50% (relative, to symptomatic)	Assumed
Percentage individuals infectious pre-onset	20%, **40%**	[9,32]
Outside infection rate	0.0001**, 0.001,** 0.005, 0.01	Assumed
Percentage of contacts traced	40%, 60%, **80%**	Assumed
Maximum number of tests	0, 5, 25, 50	Tested
Test false positive rate	0.02	[24]
Test false negative rate	0.1 (symptomatic patients) 0.5 (asymptomatic patients)	Based data from early infection stages in [23]

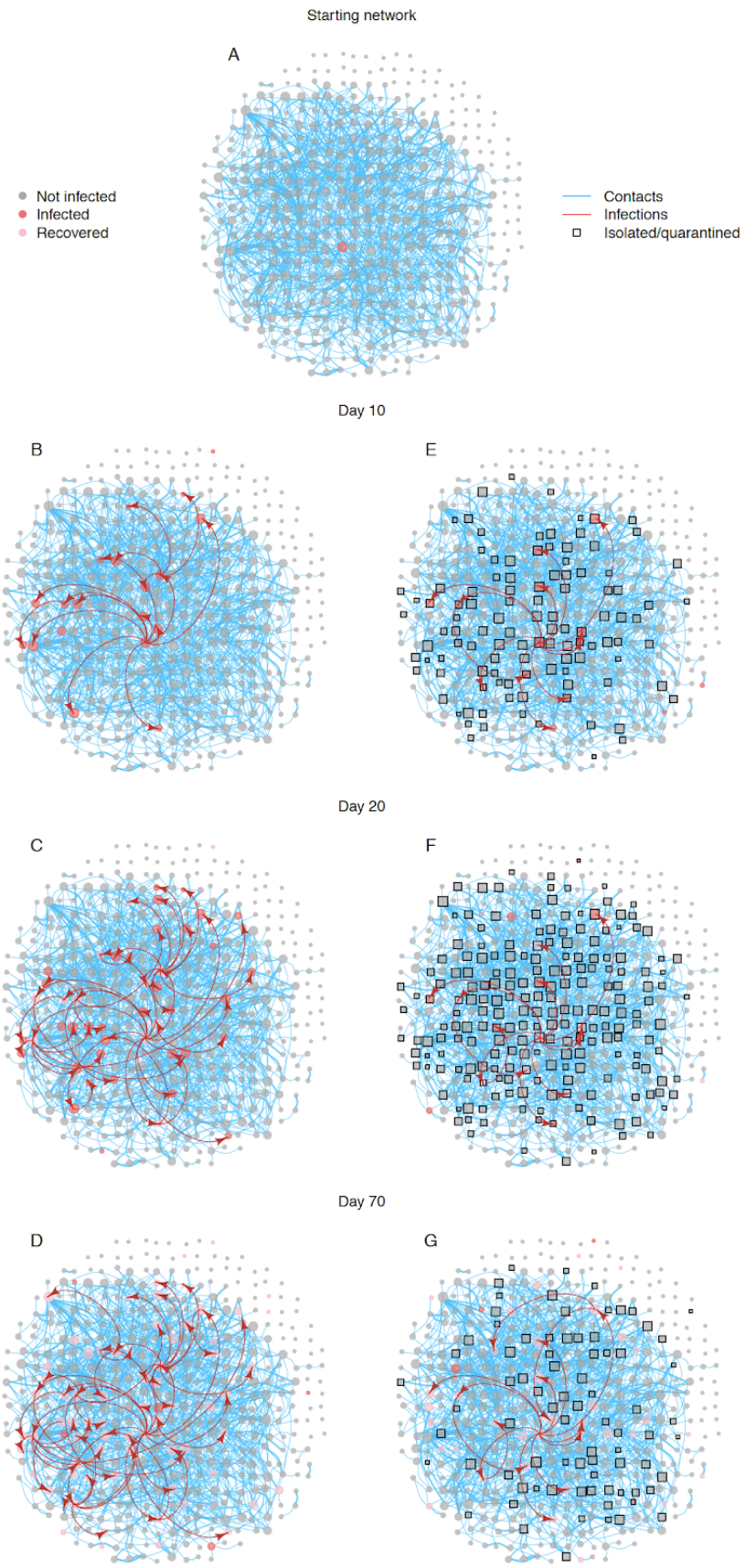

Figure 1 Illustration of the Haslemere network with epidemic simulation predictions for 'nothing' (left side) and secondary contact tracing scenarios (right side). **A** The social network of 468 individuals (grey nodes) with 1257 social links (blue edges) weighted by 1616 daily contacts (edge thickness) and a single starting infector (red). Subsequent panels show progression of the COVID-19 epidemic under the nothing (**B,C,D**) and the secondary contact tracing (**E,F,G**) scenarios. Red arrows show an infection route, and squares show isolated/quarantined individuals. See Supplementary Video 1 for animated visualisation of all scenarios, and a Shiny app is available to run individual outbreak simulations (see data sharing)

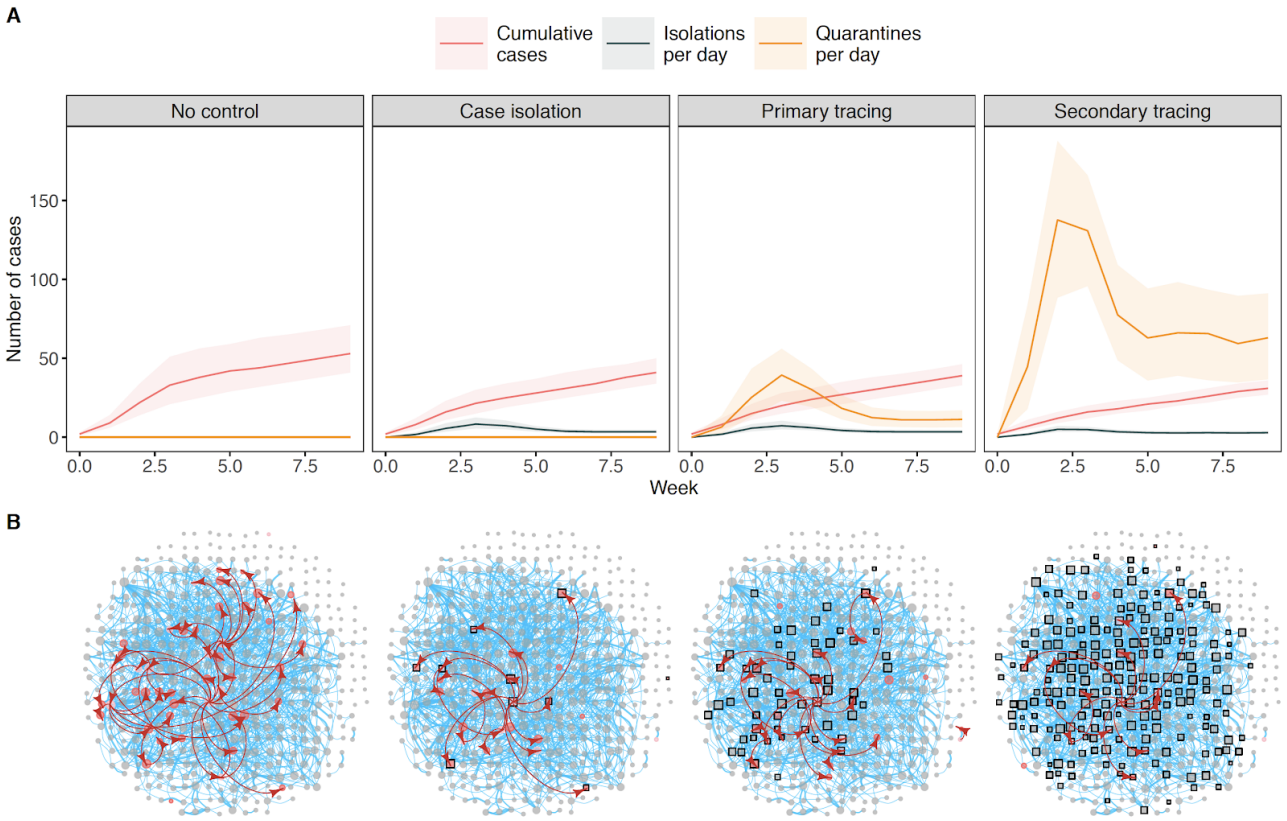

Figure 2 Epidemic model predictions of outbreak size and number of people isolated/quarantined under different non-pharmaceutical intervention scenarios in the Haslemere network. **A** cumulative number of cases, number of people isolated per day, and number of people quarantined per day under each scenario. Lines and shaded areas represent median and interquartile range from 1000 simulations. **B** Example networks from a single simulation of each scenario at day 20 of the outbreak. See figure 1 for network details.

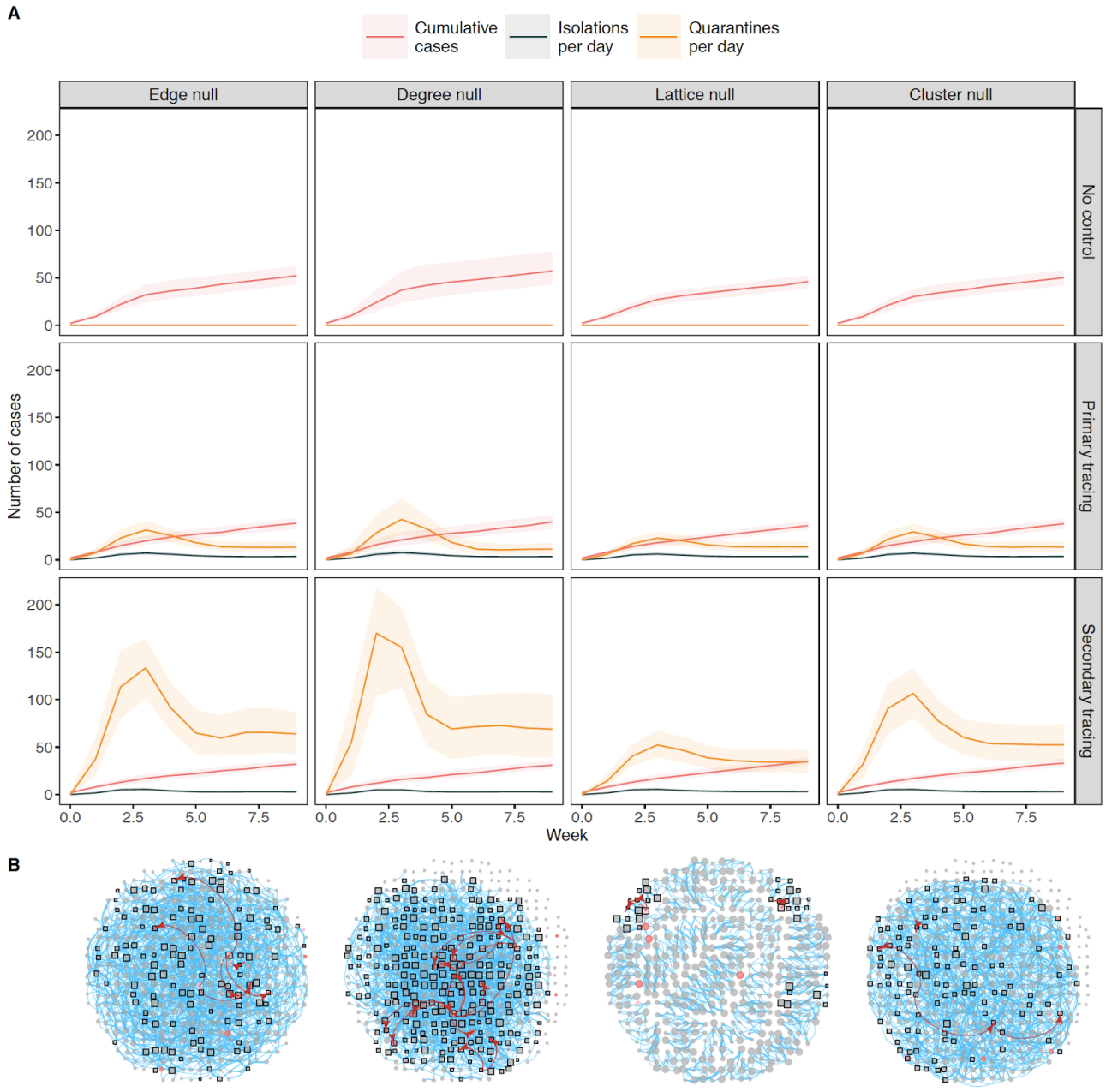

Figure 3 A Epidemic model simulations of outbreak size and number of people isolated/quarantined under different null-network permutations based on the Haslemere network (see methods for details). Lines and shaded areas represent median and interquartile range from 1000 simulations. **B** Example networks showing an infection simulation (with secondary contact tracing, after 20 days) on each null network. See Figure 1 for network details.

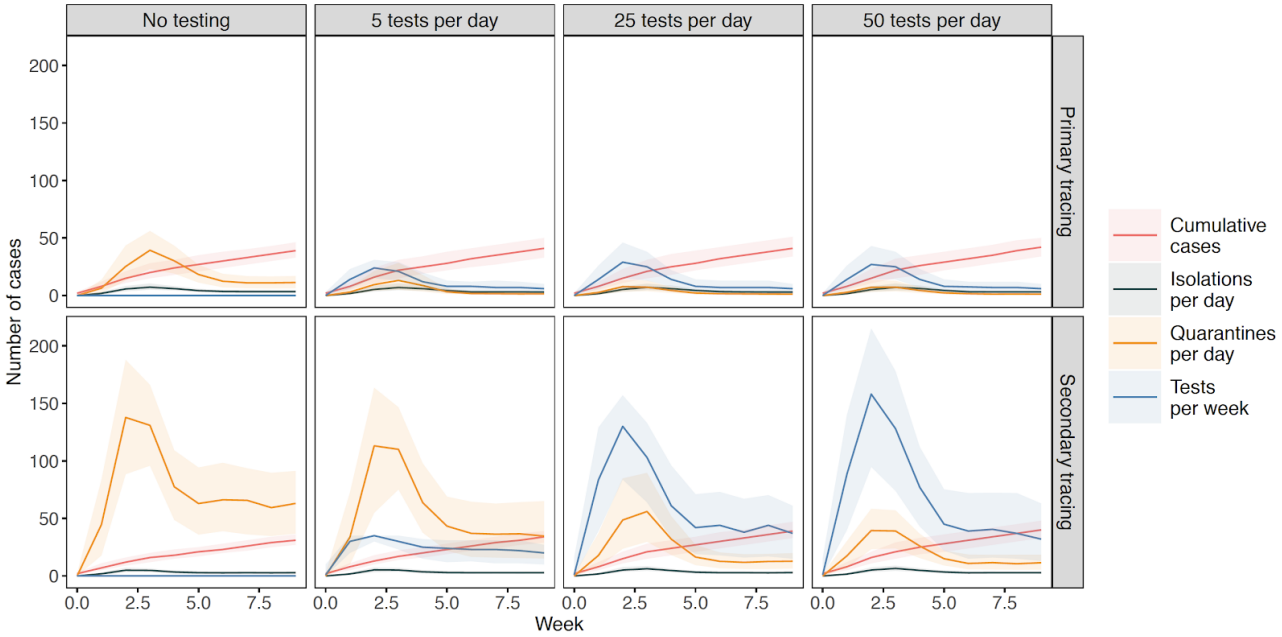

Figure 4 Epidemic model predictions of outbreak size and number of people isolated, quarantined and tested under different testing rates in the Haslemere network. Tests are plotted per week rather than per day for visualisation purposes. Lines and shaded areas represent median and interquartile range from 1000 simulations.

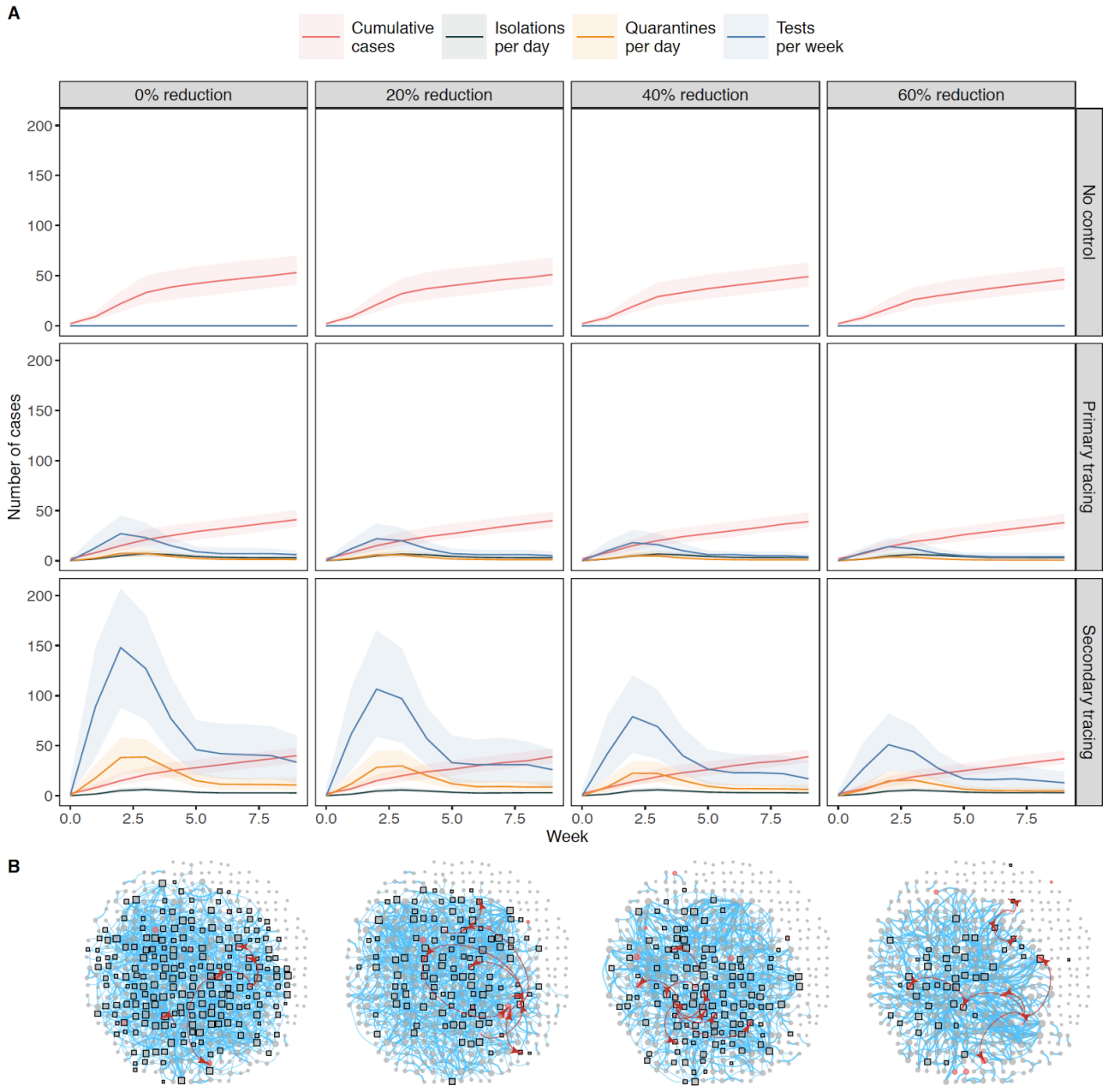

Figure 5 A Epidemic model simulations of outbreak size and number of people isolated, quarantined and tested under different levels of physical distancing in the Haslemere network. The percentage reduction refers to the number of 'weak links' reassigned within the networks to increase clustering (see methods). Tests are plotted per week rather than per day for visualisation purposes. Lines and shaded areas represent median and interquartile range from 1000 simulations. **B** Example networks showing an infection simulation (with no control, for 70 days) at each level of physical distancing. See figure 1 for network details.