LETTERS
https://doi.org/10.1038/s41591-020-1036-8

nature medicine

Check for updates

Using a real-world network to model localized COVID-19 control strategies

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

Case isolation and contact tracing can contribute to the control of COVID-19 outbreaks[1,2]. However, it remains unclear how real-world social networks could influence the effectiveness and efficiency of such approaches. To address this issue, we simulated control strategies for SARS-CoV-2 transmission in a real-world social network generated from high-resolution GPS data that were gathered in the course of a citizen-science experiment[3,4]. We found that tracing the contacts of contacts reduced the size of simulated outbreaks more than tracing of only contacts, but this strategy also resulted in almost half of the local population being quarantined at a single point in time. Testing and releasing non-infectious individuals from quarantine led to increases in outbreak size, suggesting that contact tracing and quarantine might be most effective as a 'local lockdown' strategy when contact rates are high. Finally, we estimated that combining physical distancing with contact tracing could enable epidemic control while reducing the number of quarantined individuals. Our findings suggest that targeted tracing and quarantine strategies would be most efficient when combined with other control measures such as physical distancing.

Non-pharmaceutical interventions are central to reducing SARS-CoV-2 transmission in the absence of an effective vaccine[5–8]. Such measures include case isolation, tracing and quarantining of contacts, use of personal protective equipment 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). Because of the varying economic and social costs of these interventions, there is a clear need for sustainable strategies that limit SARS-CoV-2 transmission while reducing disruption as much as possible.

Isolation of symptomatic individuals and quarantine of their contacts (for example, household members) is a common public health strategy for reducing infectious disease spread[1,2,8]. This approach has been used as part of SARS-CoV-2 control strategies globally[9]. However, the relatively high reproduction number of the SARS-CoV-2 virus in early outbreak stages[10,11], alongside likely high contribution to transmission from presymptomatic and asymptomatic individuals[12], means that manual tracing of contacts alone might not be a sufficient containment strategy under a range of outbreak scenarios[13]. 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[14].

It is possible to assess the potential effectiveness of contact tracing by simultaneously modeling disease spread and contact tracing strategies through social systems of individuals[15]. These systems are usually simulated through parameterization with simple social behaviors (for example, the distribution of the number of physical contacts per individual). Furthermore, social systems can be simulated as networks that are parameterized according to assumptions regarding different contexts (for example, with different simulated networks for households, schools and workplaces) or using estimated contact rates for different age groups[16]. However, little is known about how different types of real-world social behavior and hidden structures in real-life networks could affect both patterns of disease transmission and the efficacy of contact tracing under different scenarios[17,18]. 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.

Here we develop an epidemic model that simulates COVID-19 outbreaks across a real-world network, and we assess the impact of a range of testing and contact tracing strategies for controlling these outbreaks. We then simulate physical distancing strategies and quantify how the interaction among physical distancing, contact tracing and testing affects outbreak dynamics. A summary of the main findings, limitations and policy implications of our study is shown in Table 1.

We used a publicly available dataset on human social interactions collected specifically for modeling infectious disease dynamics as part of the BBC documentary 'Contagion! The BBC Four Pandemic' (refs. [3,4]). 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 (ref. [19]). This dataset is structurally relevant to modeling disease spread and hence holds substantial potential for understanding and controlling the spread of real-world infectious diseases[3,4]. Here we defined dyadic contacts on a day-by-day basis as at least one daily 5-min period in which the distance between the individuals was within 4 m (Methods), which gave 1,616 daily contact events and 1,257 unique social links among 468 individuals. The social network was therefore weighted by the number of days that individuals made contact. This network was strongly correlated ($r > 0.85$ in all cases) with social networks generated using different distances for defining contacts (contact ranges from 1–7 m;

[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, Cambridge, UK. [5]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [6]School of Biological Sciences, University of East Anglia, Norwich, UK. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: l.spurgin@uea.ac.uk

Table 1 | Policy summary

Background	Understanding how isolation, contact tracing and other non-pharmaceutical interventions can be combined effectively and efficiently is crucial to maintaining COVID-19 control. We developed an epidemic model that simulates COVID-19 outbreaks in a real-world network and assessed the impact of a range of testing, isolation, quarantine and contact tracing strategies for controlling new local outbreaks.
Main findings and limitations	We found that tracing and quarantining contacts of contacts was the most effective simulated measure for controlling local COVID-19 outbreaks but required large numbers of individuals to be quarantined. This strategy is similar to introducing a local lockdown. Testing and releasing quarantined individuals reduced the numbers quarantined but also reduced the effectiveness of control measures. Combining physical distancing with contact tracing resulted in reduced outbreak size, with fewer individuals required to quarantine. A major limitation of this study is that it is based on pre-COVID-19 social network data from a sample of individuals from a single small town; more data are needed to fully understand potential outbreak dynamics in other settings.
Policy implications	Our findings suggest that effective contact tracing measures could require large numbers of people in a community to be quarantined, with individual-level tracing resulting in outcomes equivalent to broad local lockdowns. Targeted tracing and quarantine strategies might be less disruptive overall when combined with other control measures such as moderate physical distancing.

Extended Data Fig. 1) and with social networks created using different time periods for weighting the dyadic contacts (Extended Data Fig. 2). As such, this social network quantification gives a representative indication of daily contact propensities within the relevant transmission range between individuals (Methods) and captures various aspects of the patterns and structure presented by different quantifications of this social system.

Example outbreaks across the Haslemere social network under different control scenarios are displayed in Fig. 1, with a full animated visualization in Supplementary Video 1 and a Shiny app available to run individual outbreak simulations (see Data availability). 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 75% (5th–95th percentiles, 72–77%) of the population infected 70 days after the first simulated infection (Fig. 2). Isolation of individuals when they became symptomatic resulted in 66% (62–69%) of the population infected, and primary contact tracing resulted in 48% (42–54%) infected. Secondary contact tracing resulted in the smallest percentage (16%, 11–22%) of the population infected after 70 days. The proportion of quarantined individuals was very high under both primary and secondary contact tracing, with a median of 43% (19–63%) of the population quarantined during the outbreak peak with secondary contact tracing (Fig. 2). Examining temporal dynamics showed that outbreak peaks typically occurred within the first 1–3 weeks following the first simulated infection and that all simulated non-pharmaceutical interventions reduced the overall size of the outbreaks as well as their growth rate (Fig. 2). The proportion 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 (26%, 8–47%) were quarantined even during the final (tenth) week of the simulations (Fig. 2). This is consistent with a large-scale simulation model of app-based contact tracing in the United Kingdom[20], which

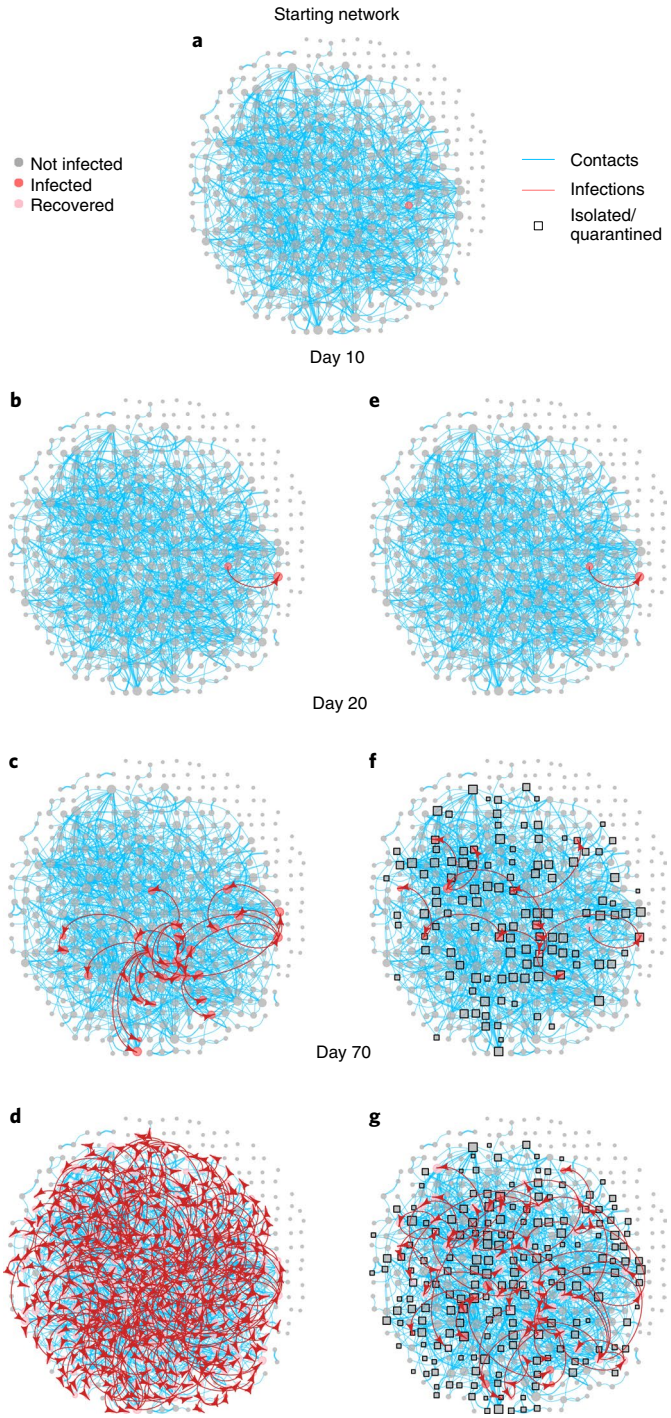


Fig. 1 | Illustration of the Haslemere network with epidemic simulation predictions. a, The social network of 468 individuals (gray nodes) with 1,257 social links (blue edges) weighted by 1,616 daily contacts (edge thickness) and a single starting infector (red). **b**–**g**, Progression of the COVID-19 epidemic under the no-intervention (**b**–**d**) and secondary contact tracing (**e**–**g**) scenarios. Red arrows indicate an infection route, and squares highlight isolated or quarantined individuals.

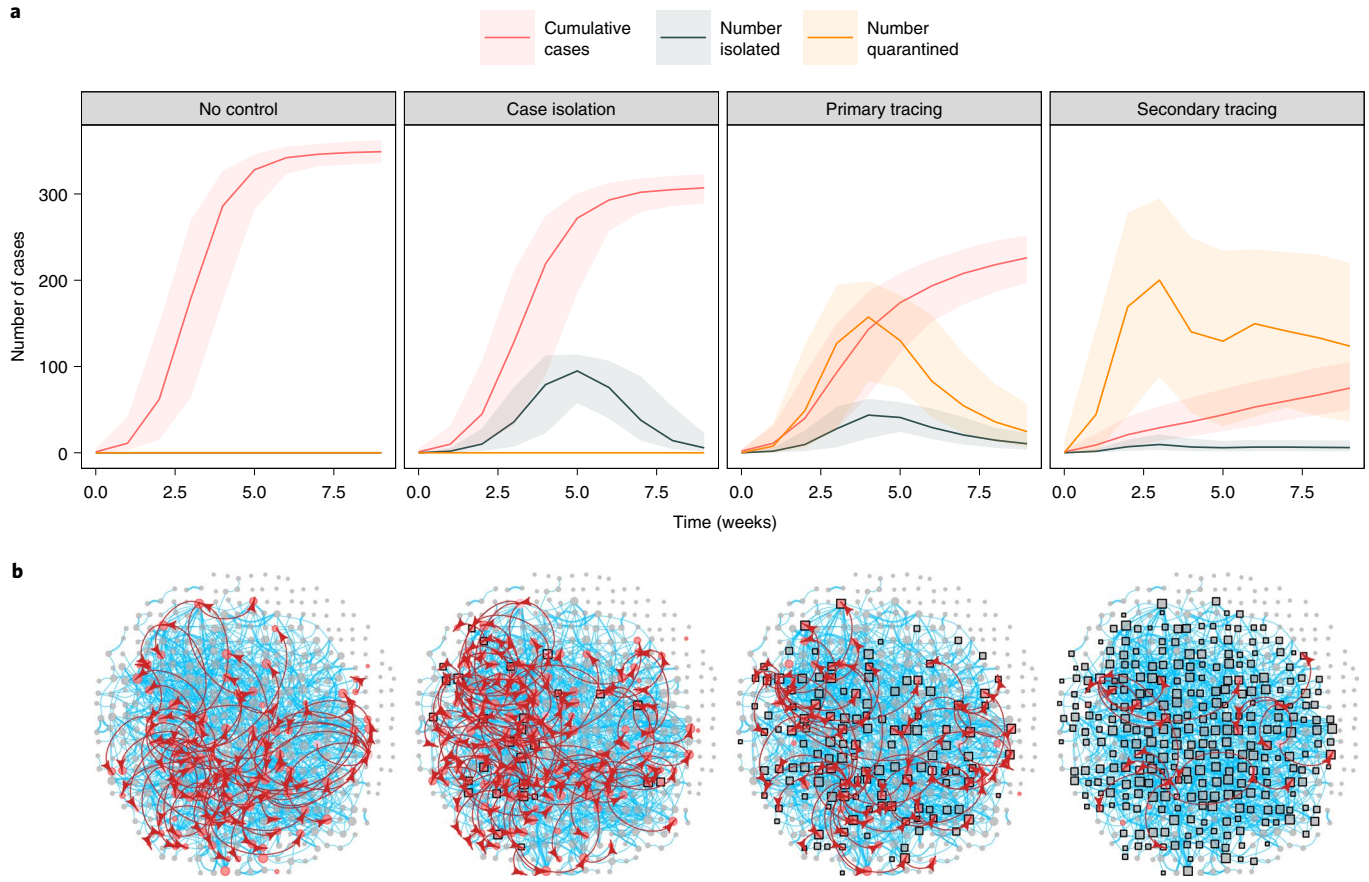


Fig. 2 | Epidemic model predictions of outbreak size and the number of people isolated or quarantined under different non-pharmaceutical intervention scenarios in the Haslemere network. a, Cumulative number of cases, number of people isolated and number of people quarantined at a given point in time under each scenario. Lines and shaded areas represent the median and 5th–95th percentiles, respectively, from 1,000 simulations. **b**, Example networks from a single simulation of each scenario at day 20 of the outbreak. See Fig. 1 for network details.

suggested that contact tracing could be highly effective, but also that it required large numbers of people to be quarantined. We assumed that 10% of contact tracing attempts were missed, which, when combined with the large number of quarantined individuals 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.

Sensitivity analysis of the efficacy of contact tracing under the epidemic model is presented in Extended Data Figs. 3–6. As expected, outbreak size decreased as the percentage of contacts traced increased in all scenarios and increased with increasing values of the reproduction number, the proportion of asymptomatic cases, the proportion of pre-onset transmission, the delay between onset/tracing and isolation/quarantine, and the number of initial cases (Extended Data Figs. 3–6). Outbreak dynamics were strongly affected by the outside infection rate across all intervention scenarios, as were the numbers of isolated and quarantined individuals (Extended Data Fig. 6). These findings suggest that, likely owing to the high levels of SARS-CoV-2 transmission from asymptomatic and presymptomatic individuals[12], contact tracing would be most effective when the proportion of traced contacts is high, when the delay from notification to quarantine is short[13] and when the number of starting cases and rate of movement into the network are low. Importantly, however, outbreak control was only achieved when there was a large number of quarantined individuals, and this was consistent across the entirety of the parameter space (Extended Data Figs. 3–6). Furthermore, increasing the network density through increasing the distance threshold for defining contacts led to broadly similar results across intervention scenarios, albeit with larger numbers of quarantined individuals required for outbreak control via contact tracing (Extended Data Fig. 7). Therefore, while more real-world networks are needed to demonstrate how well these results apply to other locations and settings, our results are robust to a range of epidemiological and network parameters.

The number of quarantined individuals can be reduced through mass testing and release of individuals who return a negative result. Conversely, if contact rates in the population are high, large-scale test-and-release strategies could provide greater opportunity for transmission and decrease the effectiveness of contact tracing. We therefore assessed how the testing and release of isolated and quarantined individuals might affect the numbers of cases and time spent in isolation and quarantine, using false-positive and false-negative rates estimated from empirical data[21,22] (Supplementary Table 1). We estimated that increasing the testing capacity (and therefore testing and releasing more quarantined individuals) led to substantial increases in outbreak size, especially under secondary contact tracing (median percentage infected, 52%; 5th–95th percentiles, 46–57%; Fig. 3a). This result occurred despite an optimistically high false-negative rate of 10%, suggesting that the increase in outbreak size with high testing rates is a result of increased transmission within the network, rather than the release of infected cases per se. Indeed, increases in outbreak size were observed even when a false-negative rate of zero was assumed. Therefore, secondary

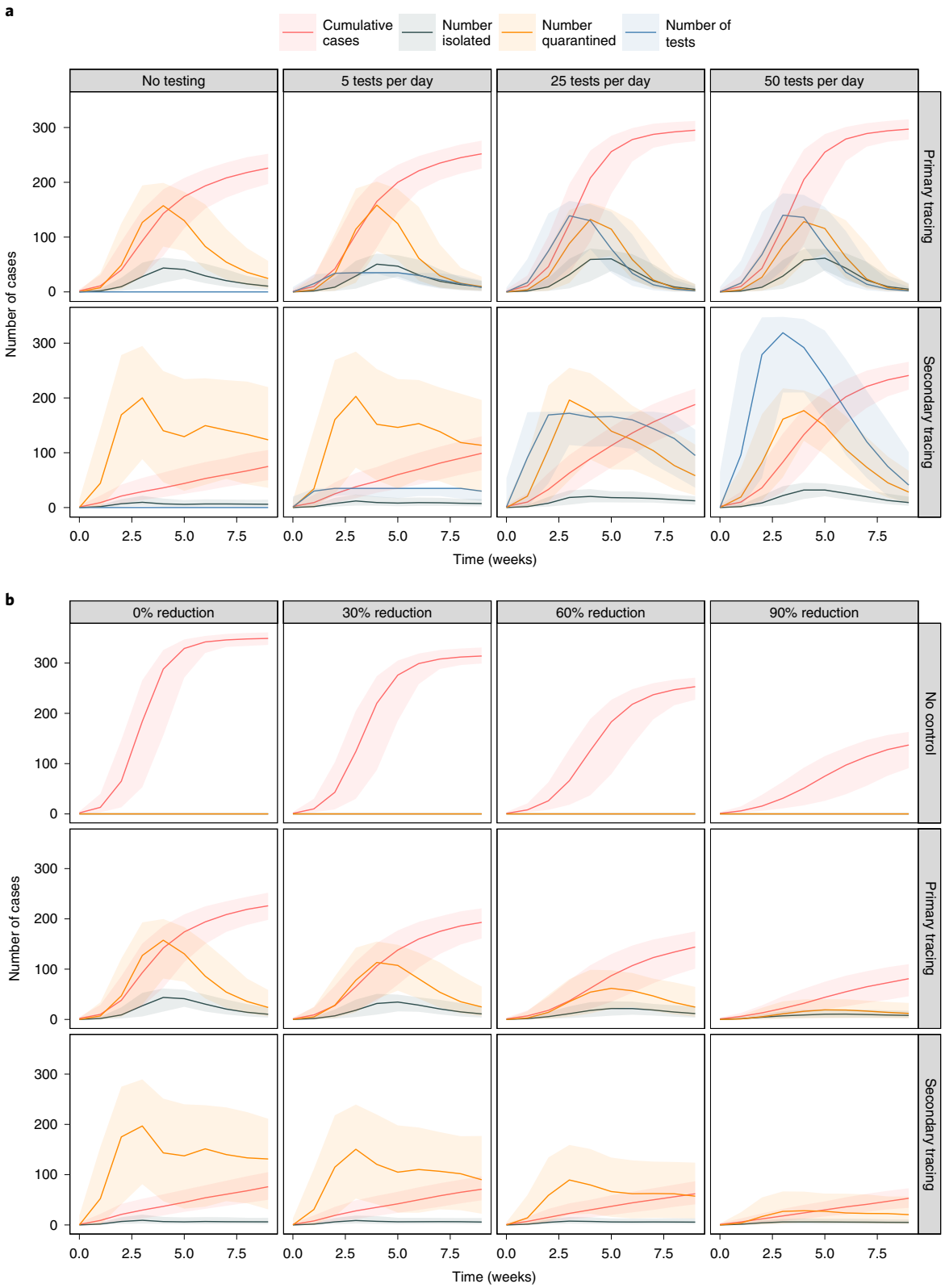


Fig. 3 | Epidemic model predictions of how testing and physical distancing affect outbreak and quarantine dynamics. a,**b**, Epidemic model simulations of outbreak size and number of people isolated and quarantined under different levels of testing (**a**) and physical distancing (**b**) in the Haslemere network. In **a**, tests are plotted per week rather than per day for visualization purposes. In **b**, the percentage reduction refers to the number of weak links removed from the networks (Methods). Lines and shaded areas represent the median and 5th–95th percentiles, respectively, from 1,000 simulations.

tracing could effectively function as a local lockdown rather than a targeted intervention strategy. High levels of testing did not lead to large reductions in the number of quarantined individuals under secondary contact tracing scenarios, and the number of tests required to reduce the proportions of quarantined individuals were large, with 68% (45–74%) of the population requiring tests in a single

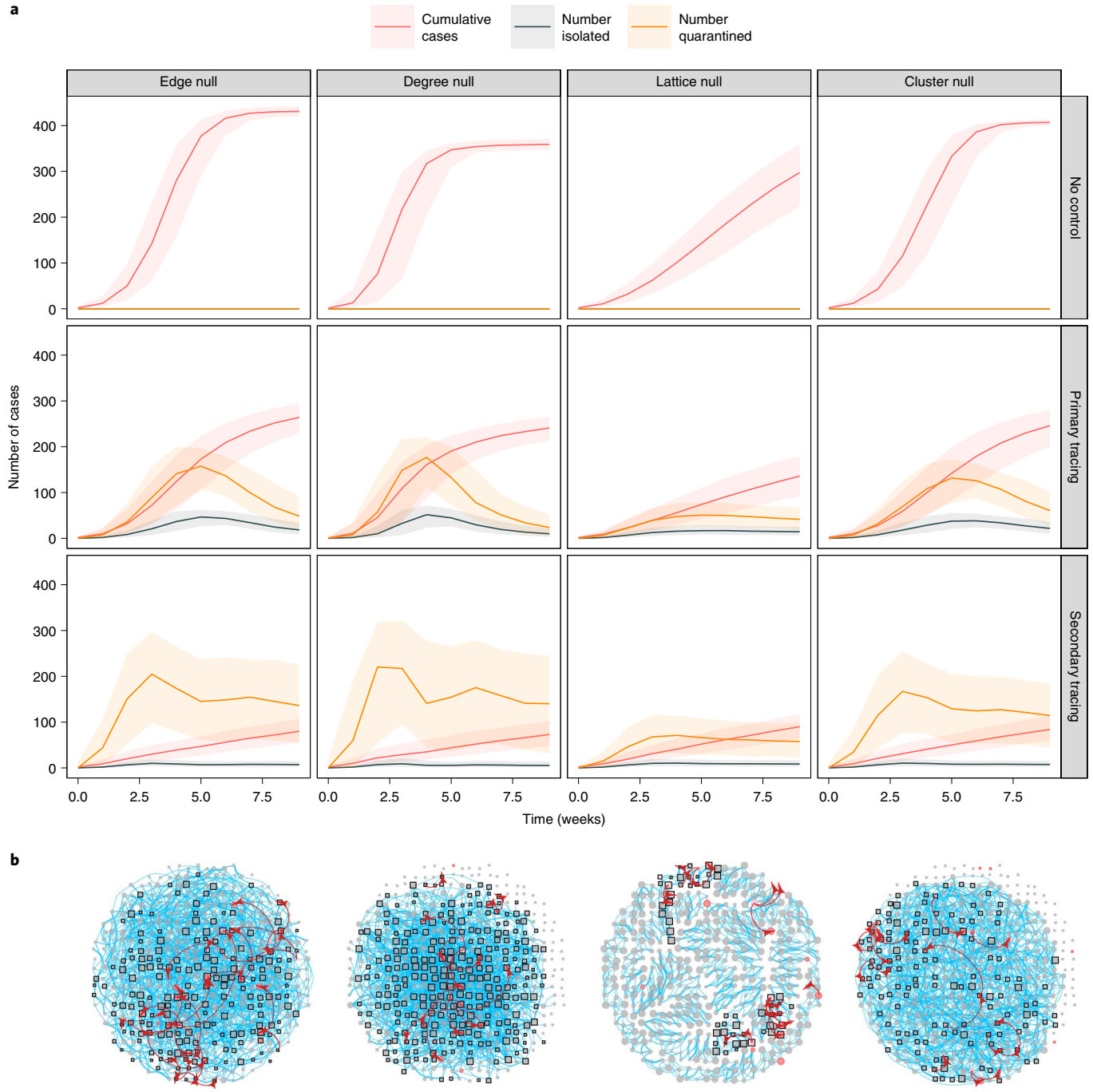


Fig. 4 | Epidemic model predictions of outbreak and quarantine dynamics under different null network permutations. a, Epidemic model simulations of outbreak size and number of people isolated and quarantined under different null network permutations based on the Haslemere network (see Methods for details). Lines and shaded areas represent the median and 5th–95th percentiles, respectively, from 1,000 simulations. **b**, Example networks showing an infection simulation (with secondary contact tracing after 20 days) on each null network. See Fig. 1 for network details.

week during outbreak peaks (Fig. 3a). We cannot be certain to what extent our results represent larger populations, but the tripartite relationship among 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.

Our model is optimistic in its assumption that individuals isolate independently of previous notifications or isolations and highly optimistic in its assumption that all traced contacts remain in quarantine for the full 14-day period. In reality, a high notification and quarantine rate could result in individuals being less likely to undertake quarantine in the future, which in turn would affect outbreak dynamics. More evidence and models to better understand these behavioral dynamics are needed to develop sustainable intervention strategies[23]. One suggested solution to reduced adherence to quarantine is through (digital) targeted quarantine requests to the individuals at highest risk of infection or to those most likely to spread to others[24]. The extent to which these interventions will be needed and how effectively they will work are not yet clear, and there are important concerns around privacy in the implementation of contact tracing strategies[25]. However, our study provides a

methodological template for network-based research into SARS-CoV-2 transmission and potential control strategies.

Combining contact tracing with other physical distancing measures could allow for outbreak control while reducing the number of people in quarantine, as well as the number of tests required. We simulated physical distancing by reducing the number of 'weak links' in the Haslemere network (Methods). We aimed to consider low to moderate levels of physical distancing, so we used a model in which the only interactions with rare contacts (those observed only on a single day) were removed. Depending on the scenario, the highest simulated levels of physical distancing led to reductions of between 28% and 61% in the number of overall cases (Fig. 3b). Importantly, increasing physical distancing was associated with a lower proportion of quarantined individuals, which was reduced to as little as 6% of the population (1–14%) during outbreak peaks under secondary contact tracing (Fig. 3b). Simulating physical distancing using an alternative approach in which removed rare contacts were reassigned to existing contacts (Methods) yielded similar results to our initial model; however, when using this approach, physical distancing led to smaller decreases in outbreak size (Extended Data Fig. 8). We do not have information on household structure within the Haslemere dataset, but our physical distancing scenario is analogous to decreasing the probability of transmission between non-household contacts. This could include physical distancing measures in public places, restrictions on large gatherings, or increased hand hygiene and use of masks outside of household settings[26]. Combining such measures with highly effective contact tracing could be a useful tool for control of SARS-CoV-2 spread. However, further work is required to determine exactly what kinds of physical distancing measures would enable effective outbreak control alongside contact tracing. Future investigations examining how the spread of the disease shapes behavioral change interventions (for example, where large outbreaks trigger more extensive physical distancing measures) and how this feedback shapes the contagion dynamics and predicted effectiveness of interventions are needed.

Network structure has substantial effects on epidemic model predictions[27,28]. We used null network models based on the Haslemere data, which maintained the same number of individuals, number of connections and weights of connections, but shuffled network architecture in different ways (Methods). The number of cases estimated using the null networks was broadly similar to the number in the real-world network, although this was substantially underestimated in a lattice-like network (Fig. 4). Importantly, the rate of quarantine varied substantially among the null networks, especially under secondary contact tracing (Fig. 4). These results demonstrate that the use of network-based simulations of SARS-CoV-2 transmission dynamics requires caution. Even if such models had precise information on the number of individuals and amount of social interaction 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. Through providing insight into how changes to network structure influence contagion dynamics, the null network simulation approach gives some indication of how contagion and associated control strategies may operate in different social environments. For example, different social structures could arise when considering particular social settings (for example, workplaces or commuting), some of which may be closer to the null networks generated here. Considering these structures will improve predictions of outbreak dynamics.

There are a number of important limitations to our study and the current availability of empirical data. Most importantly, this social network is taken from data collected in a single small town over a short period of time. We do not know to what extent the social dynamics will be applicable to larger cities and other contexts and over long periods of time. Future large-scale efforts in gathering data on dynamic fine-scale social behavior over longer periods of time (ideally over the entire contagion period) in major cities would be beneficial for assessing the relative uses of SARS-CoV-2 control strategies and for understanding how and why interventions implemented have been relatively more successful in some cities than in others[29]. Furthermore, detailed real-world data could be used to parameterize more realistic simulations of human social mixing patterns. The epidemic network-based model provided here could be applied generally to larger-scale real or simulated social networks if such data become available in the future. The Haslemere data, although rich, do 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. The limited available evidence suggests that children are less susceptible to COVID-19 than adults and may therefore play a smaller role in transmission[30]. The ability to track children will also be limited in real-world contact tracing attempts, 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 that explicitly model these limitations but lack fine-scale social tracking data[20]. 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 modeling, which may provide a promising way forward for optimizing contact tracing strategies and other non-pharmaceutical interventions.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-020-1036-8.

Received: 25 May 2020; Accepted: 27 July 2020;
Published online: 7 August 2020

References

1. Fraser, C., Riley, S., Anderson, R. M. & Ferguson, N. M. Factors that make an infectious disease outbreak controllable. *Proc. Natl Acad. Sci. USA* **101**, 6146–6151 (2004).
2. Peak, C. M., Childs, L. M., Grad, Y. H. & Buckee, C. O. Comparing nonpharmaceutical interventions for containing emerging epidemics. *Proc. Natl Acad. Sci. USA* **114**, 4023–4028 (2017).
3. 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. Preprint at *bioRxiv* https://doi.org/10.1101/479154 (2018).
4. Klepac, P., Kissler, S. & Gog, J. Contagion! The BBC Four Pandemic—the model behind the documentary. *Epidemics* **24**, 49–59 (2018).
5. Ferguson, N. et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf (2020).
6. 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).
7. Tian, H. et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* **368**, 638–642 (2020).
8. Aleta, A. et al. Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic. Preprint at *medRxiv* https://doi.org/10.1101/2020.05.06.20092841 (2020).
9. Chen, S. Coronavirus: 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 (13 April 2020).
10. Kucharski, A. J. et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* **20**, 553–558 (2020).
11. Klinkenberg, D., Fraser, C. & Heesterbeek, H. The effectiveness of contact tracing in emerging epidemics. *PLoS ONE* **1**, e12 (2006).
12. He, X. et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).

13. Hellewell, J. et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob. Health* **8**, e488–e496 (2020).
14. Ferretti, L. et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).
15. Eames, K. T. D. & Keeling, M. J. Contact tracing and disease control. *Proc. Biol. Sci.* **270**, 2565–2571 (2003).
16. Del Valle, S. Y., Hyman, J. M., Hethcote, H. W. & Eubank, S. G. Mixing patterns between age groups in social networks. *Soc. Netw.* **29**, 539–554 (2007).
17. Kiss, I. Z., Green, D. M. & Kao, R. R. Disease contact tracing in random and clustered networks. *Proc. Biol. Sci.* **272**, 1407–1414 (2005).
18. 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).
19. Coronavirus patient first to be infected in UK. *BBC News* https://www.bbc.co.uk/news/uk-51683428 (29 February 2020).
20. 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).
21. Chau, N. V. V. et al. The natural history and transmission potential of asymptomatic severe acute respiratory syndrome coronavirus 2 infection. *Clin, Infect, Dis.* https://doi.org/10.1093/cid/ciaa711 (2020).
22. Cohen, A. N. & Kessel, B. False positives in reverse transcription PCR testing for SARS-CoV-2. Preprint at *medRxiv* https://doi.org/10.1101/2020.04.26.20080911 (2020).
23. West, R., Michie, S., Rubin, G. J. & Amlôt, R. Applying principles of behaviour change to reduce SARS-CoV-2 transmission. *Nat. Hum. Behav.* **4**, 451–459 (2020).
24. McCall, B. Shut down and reboot—preparing to minimise infection in a post-COVID-19 era. *Lancet Digit. Health* **2**, e293–e294 (2020).
25. Zastrow, M. South Korea is reporting intimate details of COVID-19 cases: has it helped? *Nature* https://doi.org/10.1038/d41586-020-00740-y (2020).
26. Ma, Q.-X. et al. Potential utilities of mask-wearing and instant hand hygiene for fighting SARS-CoV-2. *J. Med. Virol.* https://doi.org/10.1002/jmv.25805 (2020).
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. Cohen, J. & Kupferschmidt, K. Countries test tactics in 'war' against COVID-19. *Science* **367**, 1287–1288 (2020).
30. Davies, N. G. et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* https://doi.org/10.1038/s41591-020-0962-9 (2020).

CMMID COVID-19 Working Group

Mark Jit[3], Katherine E. Atkins[3], Samuel Clifford[3], C. Julian Villabona-Arenas[3], Sophie R. Meakin[3], Charlie Diamond[3], Nikos I. Bosse[3], James D. Munday[3], Kiesha Prem[3], Anna M. Foss[3], Emily S. Nightingale[3], Kevin van Zandvoort[3], Nicholas G. Davies[3], Hamish P. Gibbs[3], Graham Medley[3], Amy Gimma[3], Stefan Flasche[3], David Simons[3], Megan Auzenbergs[3], Timothy W. Russell[3], Billy J. Quilty[3], Eleanor M. Rees[3], Quentin J. Leclerc[3], W. John Edmunds[3], Sebastian Funk[3], Rein M. G. J. Houben[3], Gwenan M. Knight[3], Sam Abbott[3], Fiona Yueqian Sun[3], Rachel Lowe[3], Damien C. Tully[3], Simon R. Procter[3], Christopher I. Jarvis[3], Akira Endo[3], Kathleen O'Reilly[3], Jon C. Emery[3], Thibaut Jombart[3], Alicia Rosello[3], Arminder K. Deol[3], Matthew Quaife[3], Stéphane Hué[3], Yang Liu[3], Rosalind M. Eggo[3] and Carl A. B. Pearson[3]

Methods

Ethics statement. Information was provided and consent was obtained from all participants in the study before the app recorded any data. The study was approved by the London School of Hygiene & Tropical Medicine Observational Research Ethics Committee (ref. 14400).

Social tracking data. The Haslemere dataset was generated and described as part of previous work, which gives a detailed description of the characteristics of this dataset and town[3,4]. Briefly, the data were collected during the 2017–2018 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[3,4]. Of the 1,272 individuals in Haslemere who downloaded the app, 468 individuals had sufficient data points at a resolution of 1 m over three full days within the focal area for further analysis[3]. All 468 focal individuals were known to have spent >6 h within the area bounded by 51.0132° N, 0.7731° W (southwest) and 51.1195° N, 0.6432° W (northeast) (within postcode GU27), but the dataset used here comprises deidentified proximity data made available as pairwise distances (~1-m resolution) at 5-min intervals (excluding the period from 11 p.m. to 7 a.m.)[3].

Social network construction. In our primary analysis, we defined social contacts as events when the average pairwise distances between individuals within a 5-min time interval (calculated using the Haversine formula for great-circle geographic distance[3]) were 4 m or less. By doing so, we aimed to capture the majority of relevant face-to-face contacts (that is, those that might result in transmission) over 5-min periods, particularly given the 1-m potential error[3] on the tracking measurement during these short time intervals. Furthermore, this threshold of 4 m 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 (thresholds of 1 m to 7 m). We first measured the correlation of the network structure (that is, 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 passed through an individual) and eigenvector centrality (a measure that accounts for both a node's centrality and that of its neighbors) across the seven networks.

The Haslemere data are a temporal dataset spanning three full days. While the epidemic model we use is dynamic, the contagion process of COVID-19 operates over a longer time period than 3 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, which would require contact data over a much longer period of time. In the primary analysis, we weighted the edges by the number of unique days a dyad was observed together (but see the 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, with 0 corresponding to no contact, 1 corresponding to 'weak links' observed on the minority of days (one-third), 2 corresponding to 'moderate links' observed on the majority of days (two-thirds) and 3 corresponding to '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. For sensitivity analysis, we also created other weightings for this network and examined the correlation in dyadic social association scores (using Mantel tests) with our primary weighting method. Specifically, for the sensitivity analysis, we used edges specified as (1) a binary (that is, unweighted) network across all days, (2) a raw (and ranked) count of the 5-min intervals in contact, (3) a transformed weighted count (edge weight transformed as $1-e^{\text{interval count}}$, which approximates a scenario where infection risk increases with contact time but reaches 95% saturation after ~15 min of contact within dyads) and (4) a simple ratio index (SRI) weighting that corrects for observation number as SRI score[31]. The SRI score for any two individuals (that is, A and B) is calculated as

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

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[32] to understand the network properties that shape predictions of COVID-19 spread under different control scenarios. Null networks can also show how contagion may depend on the arrangement of social ties, how it 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 (Extended Data Fig. 9) with 1,000 networks generated under each of these. All of the null network scenarios kept the same number of nodes, number of edges and weights of these edges as the Haslemere network but were generated under the following nulls: (1) the 'edge null' scenario (Extended Data Fig. 9a) considered random social associates, allowing the edges of the network to be randomly allocated among all nodes; (2) the 'degree null' scenario (Extended Data Fig. 9b) considered individual differences in sociality but assigned random social links for dyads, so randomly swapped the edges between nodes but maintained the degree distribution of the real network (this was therefore even more conservative than a power-law network simulation aiming to match real differences in sociality); (3) the 'lattice null' scenario (Extended Data Fig. 9c) considered triadic and tight clique associations, so created a ring-like lattice structure through assigning all edges into a ring lattice where individuals were connected to their direct neighbors and their second- and third-order neighbors (that is, six links per individual), from which excess links were then randomly removed (until the observed number of edges was reached); and (4) the 'cluster null' scenario (Extended Data Fig. 9d) considered the observed level of clustering, so created a ring lattice structure as described above but only between individuals observed as connected (at least one social link) in the real network, added remaining links (sampled from fourth-order neighbors) and then rewired the edges until the real-world global clustering was observed (~20% rewiring; Extended Data Fig. 9d). These conservative (and informed) null models allowed connections to be arranged differently within the network but maintained the exact same number of individuals, number of social connections and weights of these social connections at each simulation.

Epidemic model. By building on the epidemiological structure of a previous branching-process model[13], we developed a full epidemic model to simulate COVID-19 dynamics across the Haslemere network. Full model parameters are given in Supplementary Table 1. For a given network of individuals, an outbreak is seeded by randomly infecting a given number of individuals (default = 1). The model then moves through daily time steps, with opportunities for infection on each day. All newly infected individuals are assigned an 'onset time' drawn from a Weibull distribution (mean = 5.8 days) that determines the point of symptom onset (for symptomatic individuals) and the point at which infectiousness is highest (for all individuals)[12]. Each individual is then simultaneously assigned asymptomatic status (whether they will develop symptoms at their onset time) and presymptomatic status (whether they will infect others before their assigned onset time), drawn from Bernoulli distributions with defined probabilities (defaults = 0.4 and 0.2, respectively; Supplementary Table 1). At the start of each day, individuals are assigned a status of susceptible, infectious or recovered (which includes deaths) on the basis of their exposure time, onset time and recovery time (calculated as the onset time plus 7 days) and are isolated or quarantined on the basis of their isolation or quarantine time. The model simulates infection dynamics over 70 days.

Possible infectors are all non-isolated and non-quarantined infectious individuals. Each day, all non-isolated, non-quarantined 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

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

where t is the number of days since 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 (Supplementary Table 1) and I_{ei} is the weighting of the edge in the network (that is, the number of days observed together) between the infector and susceptible individual. The probability density function $f\left(u;\ \mu_i, \alpha_{p_i}, \omega_{p_i}\right)$ corresponds to the generation time, which is drawn from a skewed normal distribution (see ref. [13] for details). Briefly, this uses the infector's onset time as the location parameter μ_i, while the slant parameter α_{pi} and the scale parameter ω_{pi} both vary according to the infector's presymptomatic transmission status (Supplementary Table 1). This enabled us to simulate a predefined rate of presymptomatic transmission while retaining a correlation structure between onset time and infectiousness, avoiding a scenario in which a large number of individuals were highly infectious on the first day of exposure (see Supplementary Table 1 and Data availability for more details).

When using this transmission rate, the probability of infection within a susceptible–infectious pair of individuals t days after the infector's exposure time is then modeled as

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

Note that the change in status from 'infectious' to 'recovered' at 7 days after symptom onset does not affect infection dynamics (as transmission rate ≈ 0 7 days after onset time in our model) but is instead used for contact tracing purposes. 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 1,000

trial simulations from a random single starting infector and quantified (1) the mean number of secondary infections from this case and (2) the time at which each secondary case was infected. We multiplied the rate of infection by a scaling parameter to obtain a baseline R_0 of 2.8, although we also performed sensitivity analysis (Supplementary Table 1). The mean generation time using this method was 6.3 days (median = 6 days). These basic parameters correspond closely to published estimates[12,33].

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 (0.3–0.9 tested; Supplementary 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 on the basis of their infection status, their symptomatic status, whether they have been traced and the control scenario. We considered four control scenarios: (1) no control, where no individuals are isolated or quarantined; (2) case isolation, where individuals isolate upon symptom onset after a delay period; (3) 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 (4) secondary contact tracing, as in scenario (3) 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 on the 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 (that is, the probability that an infectious individual will test negative) of 0.1 (ref. [21]) and a false-positive rate (that is, the probability that a susceptible individual will test positive) of 0.02 (ref. [22]). Individuals who test negative are immediately released from isolation or quarantine.

A set of default parameters was 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 (90%) of contacts traced within this tracked population (default parameters highlighted in bold in Supplementary Table 1). We assumed that the probability of tracing was constant over time and therefore independent of previous isolation and 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 (Supplementary Table 1). To examine how infection dynamics were affected by network structure, we ran epidemic simulations on each of the null networks described above. We also ran simulations on networks generated using higher distance thresholds (7 m and 16 m) for defining a contact. These networks were 20% and 100% more dense, respectively, and therefore provide an estimate of the robustness of our simulations to missing contacts.

We ran each simulation for 70 days, at which point the majority of new infections came from outside the network, with all scenarios replicated 1,000 times. With the null networks and physical distancing simulations, we ran one replicate simulation on each of 1,000 simulated networks. 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 calculated the number of people contained on each day of the outbreak and averaged this over weeks to obtain weekly changes in the daily rates of isolation and quarantine.

Physical distancing simulations. We simulated a population-level physical distancing effort, in which a given proportion of the weak links were removed (edges observed on only a single day; Extended Data Fig. 10a–d). This is akin to a simple situation in which individuals reduce their non-regular contacts (for example, with people outside of their household or other frequently visited settings such as workplaces). As further supplementary analysis, we also carried out a more complex physical distancing simulation in which the weak links that were removed were randomly reassigned to existing contacts (Extended Data Fig. 10e–g). This represents a scenario where individuals reduce their non-regular contacts but spend more time with regular contacts.

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

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

This study used the raw data previously published in Kissler et al.[3], which are available to download at https://github.com/skissler/haslemere. The summarized network data used here are publicly available with the code.

Code availability

The code and data used to produce the simulations are available as an R package at https://github.com/biouea/covidhm. A Shiny app that runs individual outbreak simulations is available at https://biouea.shinyapps.io/covidhm_shiny/.

References

31. Cairns, S. J. & Schwager, S. J. A comparison of association indices. *Anim. Behav.* **35**, 1454–1469 (1987).
32. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).
33. Davies, N. G. et al. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. *Lancet Public Health* **5**, e375–e385 (2020).

Acknowledgements

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

Author contributions

J.A.F., A.J.K. and L.G.S. conceived the study. J.A.F. carried out the social network analysis, with input from P.K., S.K., A.J.K. and L.G.S. L.G.S. built the epidemic network model with input from J.A.F., J.H., S.K., P.K. and A.J.K. J.A.F. and L.G.S. wrote the first draft of the manuscript. All authors interpreted the results, contributed to writing and approved the final version for submission.

Competing interests

The authors declare no competing interests.

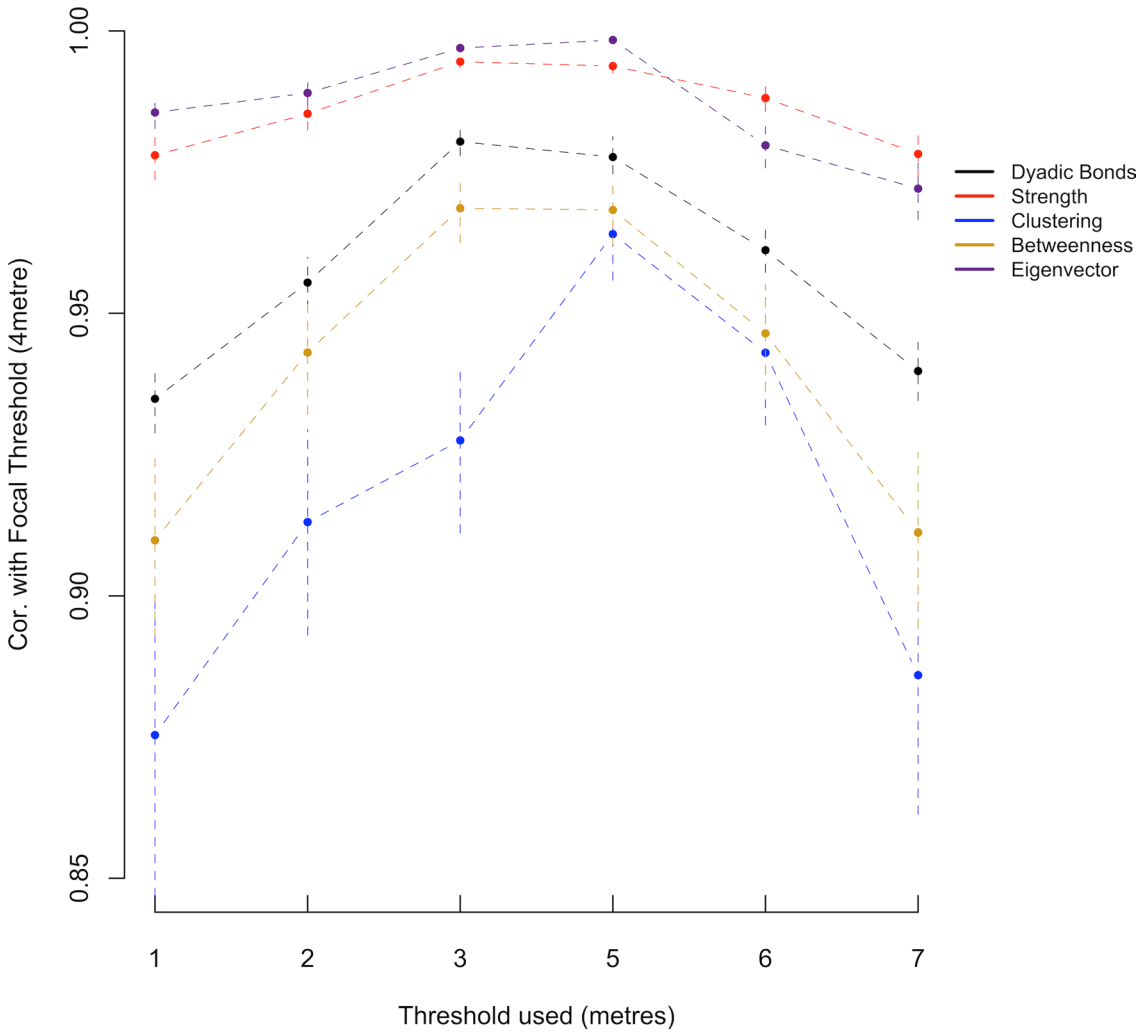
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-020-1036-8.

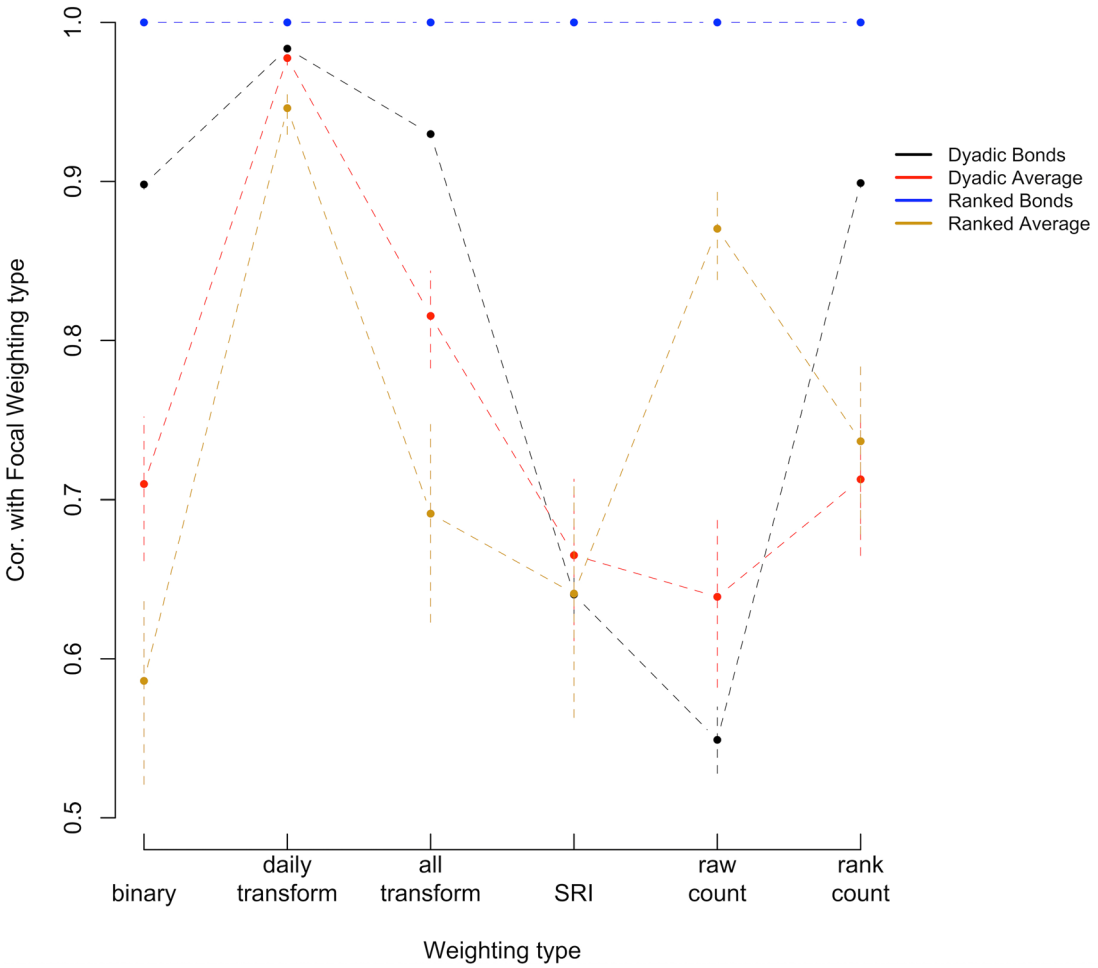
Supplementary information is available for this paper at https://doi.org/10.1038/s41591-020-1036-8.

Correspondence and requests for materials should be addressed to L.G.S.

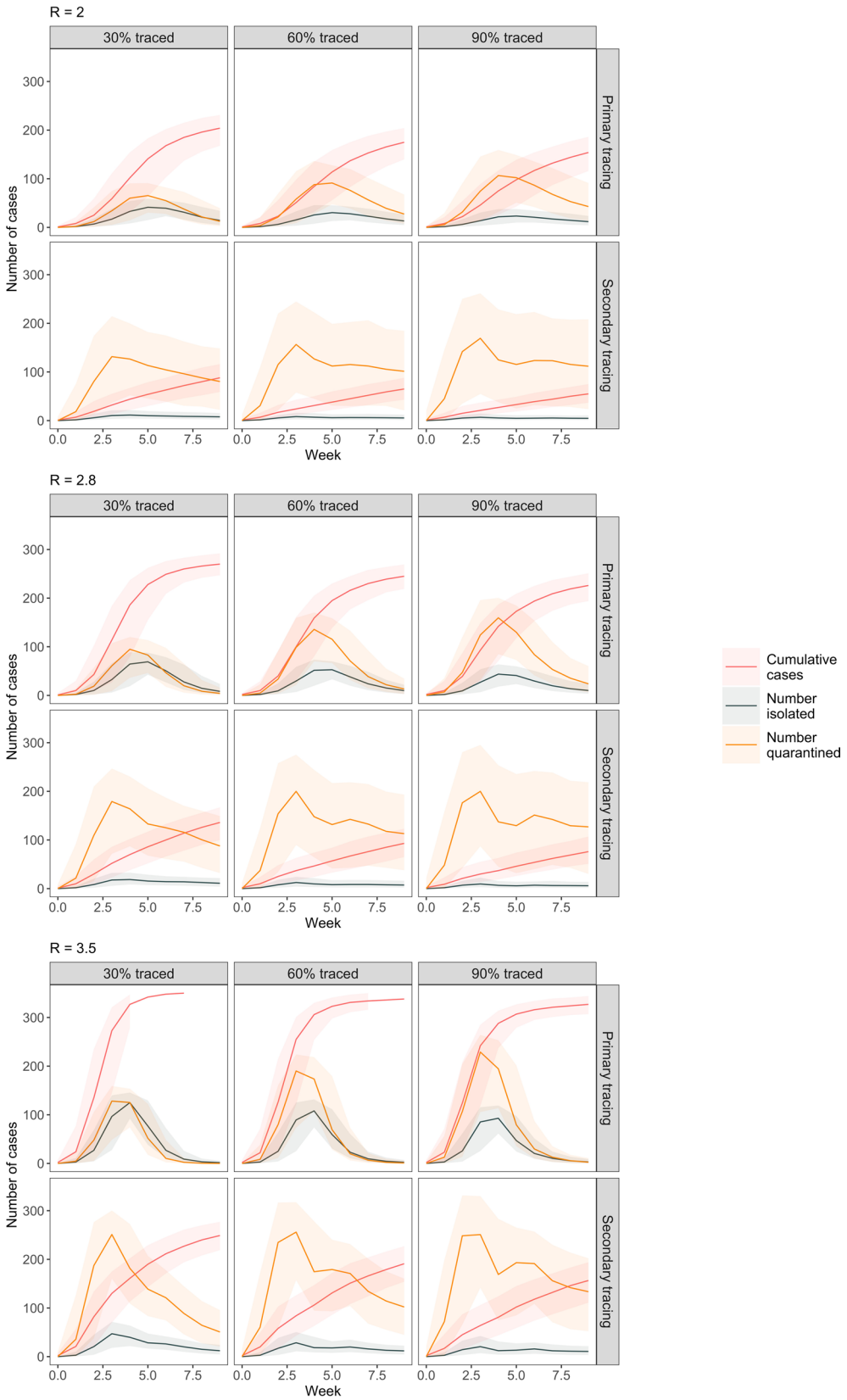
Peer review information Jennifer Sargent was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.



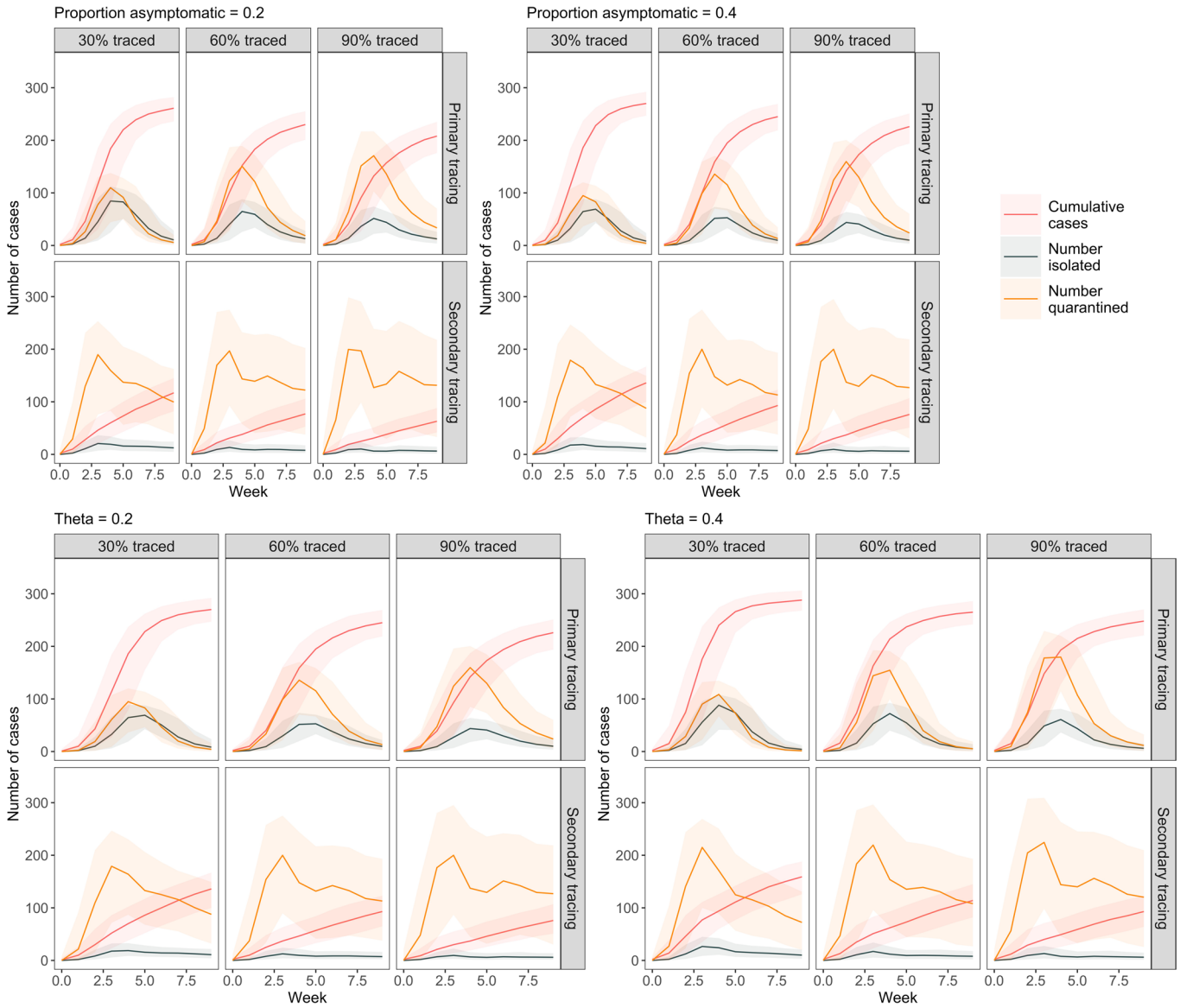
Extended Data Fig. 1 | Threshold distances and network properties. Similarity between the focal 4 m network (weighted by number of days seen together) with the other potential threshold distances defined using different average distances within the 5 min intervals (1–7 m thresholds). We considered the correlation in dyadic social associations scores between 468 individuals using Mantel tests, and examined the correlation in 468 individuals' network metrics in terms of 'weighted degree' (number of contacts with others), 'clustering coefficient' (propensity for associates to also be associated with one another), 'betweenness' (propensity to bridge the network), and 'eigenvector centrality' (the social centrality of associates). The points show the correlation coefficient and the vertical lines show the 95% confidence intervals (derived from bootstrapping the data 10,000 times) around this estimate.



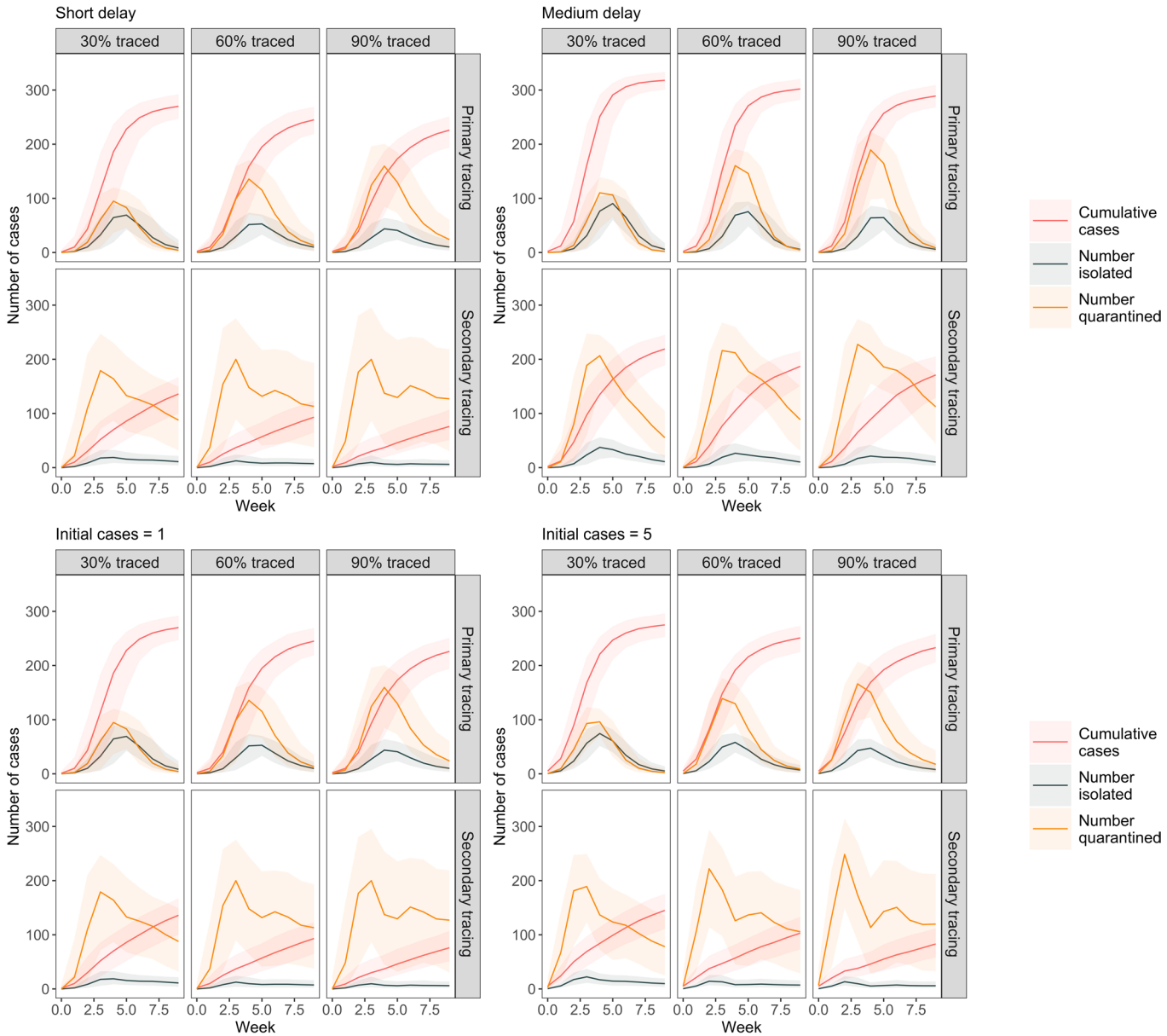
Extended Data Fig. 2 | Network weighting types and network properties. Similarity between the focal weighted network (weighted by number of days seen together) with the other potential weighting options specified here as 'binary' (whether or not individuals had social contact over the three day period), 'daily transform' (sum of $1 - e^{contacts}$ calculated for each day, where contacts = 5 min interval together each day), 'all transform' ($1 - e^{contacts}$ where contacts = 5 min interval together over all of the time period), 'SRI' (the 'Simple Ratio Index' that is using the number of 5 min intervals each dyad was seen together but correcting for the amount of 5 min intervals both members of the dyad were seen in total), 'raw count' (the number of 5 min intervals each dyad was seen together), 'rank count' (the ranked number of 5 min intervals each dyad was seen together). We calculated the network correlations (from 468 individuals) in dyadic social associations scores using two-sided Mantel tests (where 'Dyadic bonds' shows Pearson correlation and 'Ranked bonds' shows Spearman correlation), as well as the two-sided correlation in the 468 individuals' network metrics in terms of their average bond strength to all those they held an edge to (where 'Dyadic Average' shows Pearson correlation and 'Ranked Average' shows Spearman correlation). The points show the correlation coefficient and the vertical lines show the 95% confidence intervals (derived from bootstrapping the data 10,000 times) around this estimate.



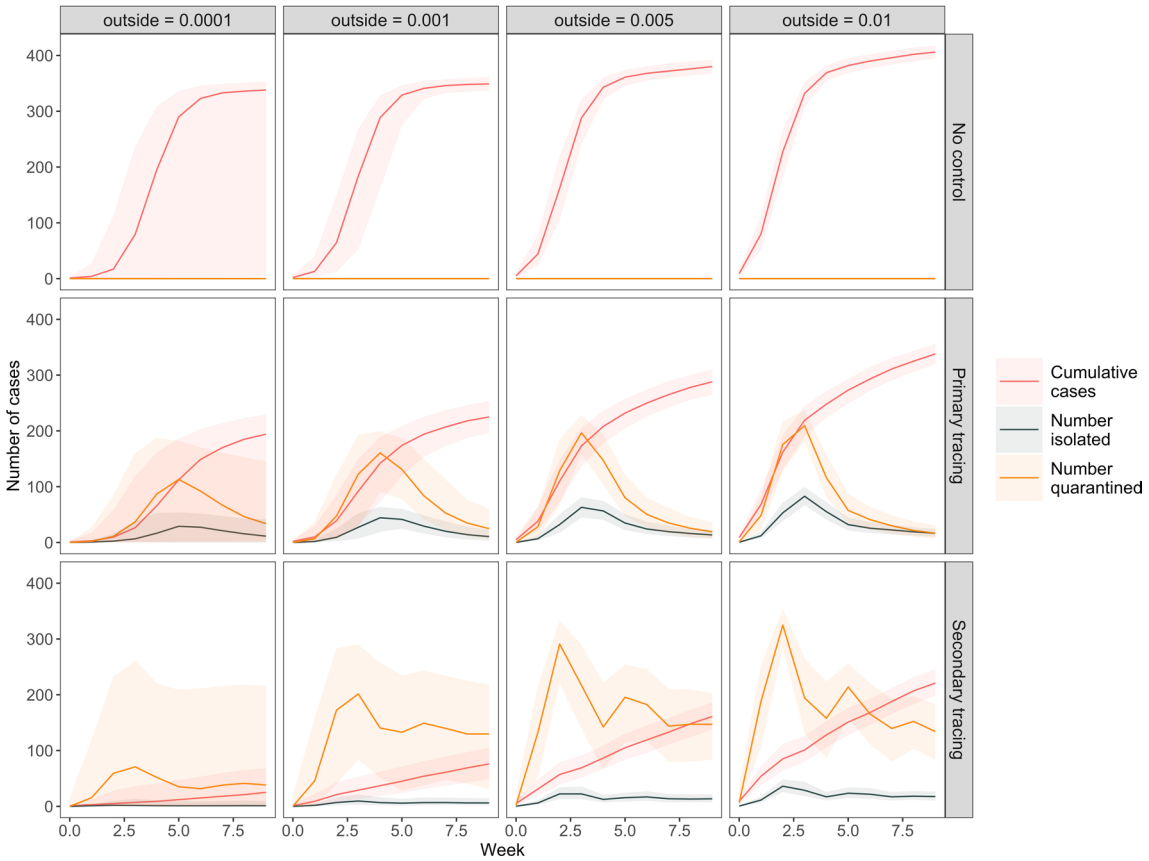
Extended Data Fig. 3 | Outbreak dynamics and and the basic reproduction number. Epidemic model predictions of outbreak size and number of people isolated/quarantined in relation to the basic reproduction number R. Lines and shaded areas represent median and 5th-95th percentiles from 1000 simulations.



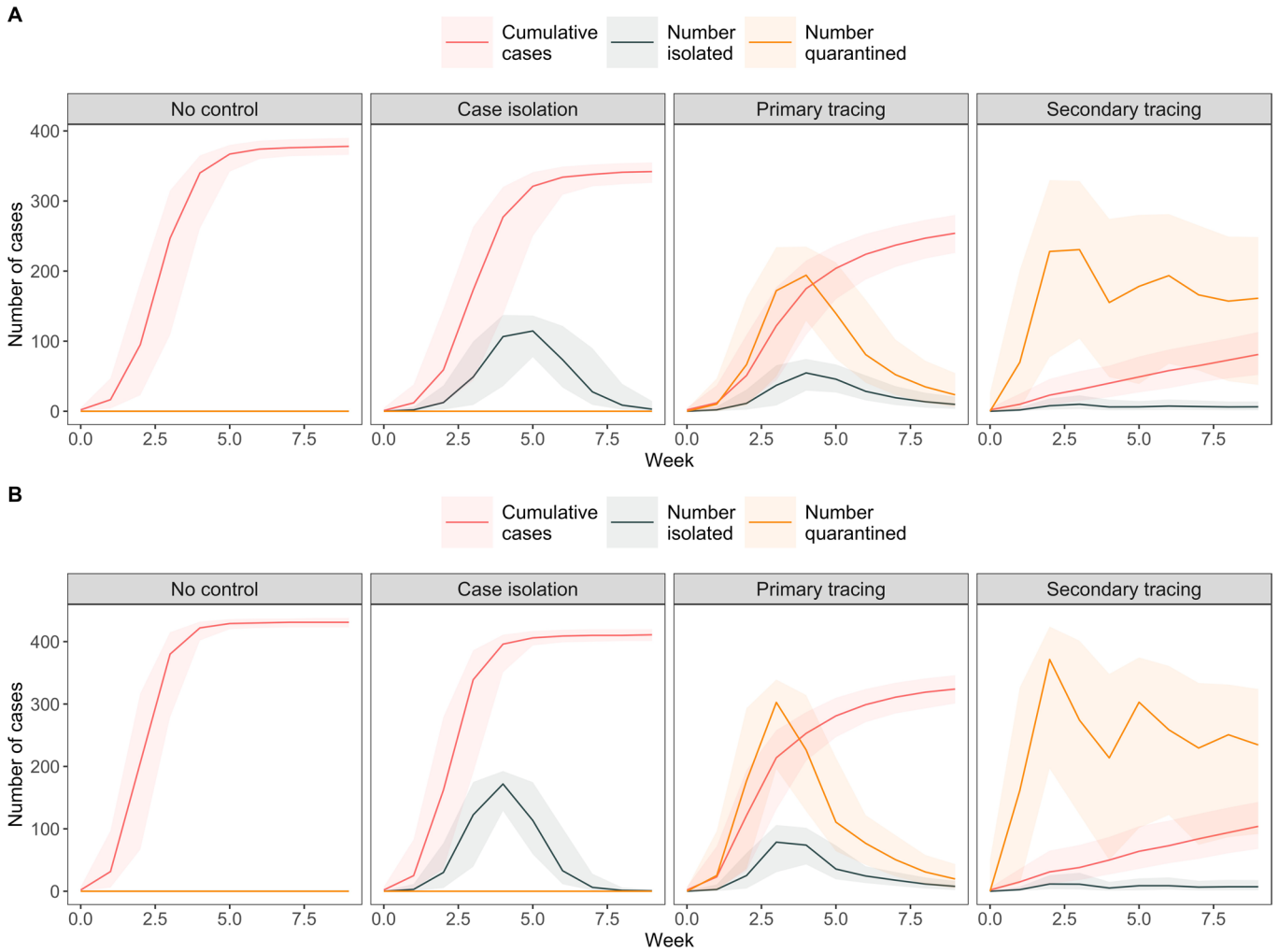
Extended Data Fig. 4 | Outbreak dynamics and asymptmocatic and presymptomatic transmission. Epidemic model predictions of outbreak size and number of people isolated/quarantined in relation to the proportion of asymptomatic cases, and the rate of presymptomatic transmission (theta). Lines and shaded areas represent median and 5th-95th percentiles from 1000 simulations.



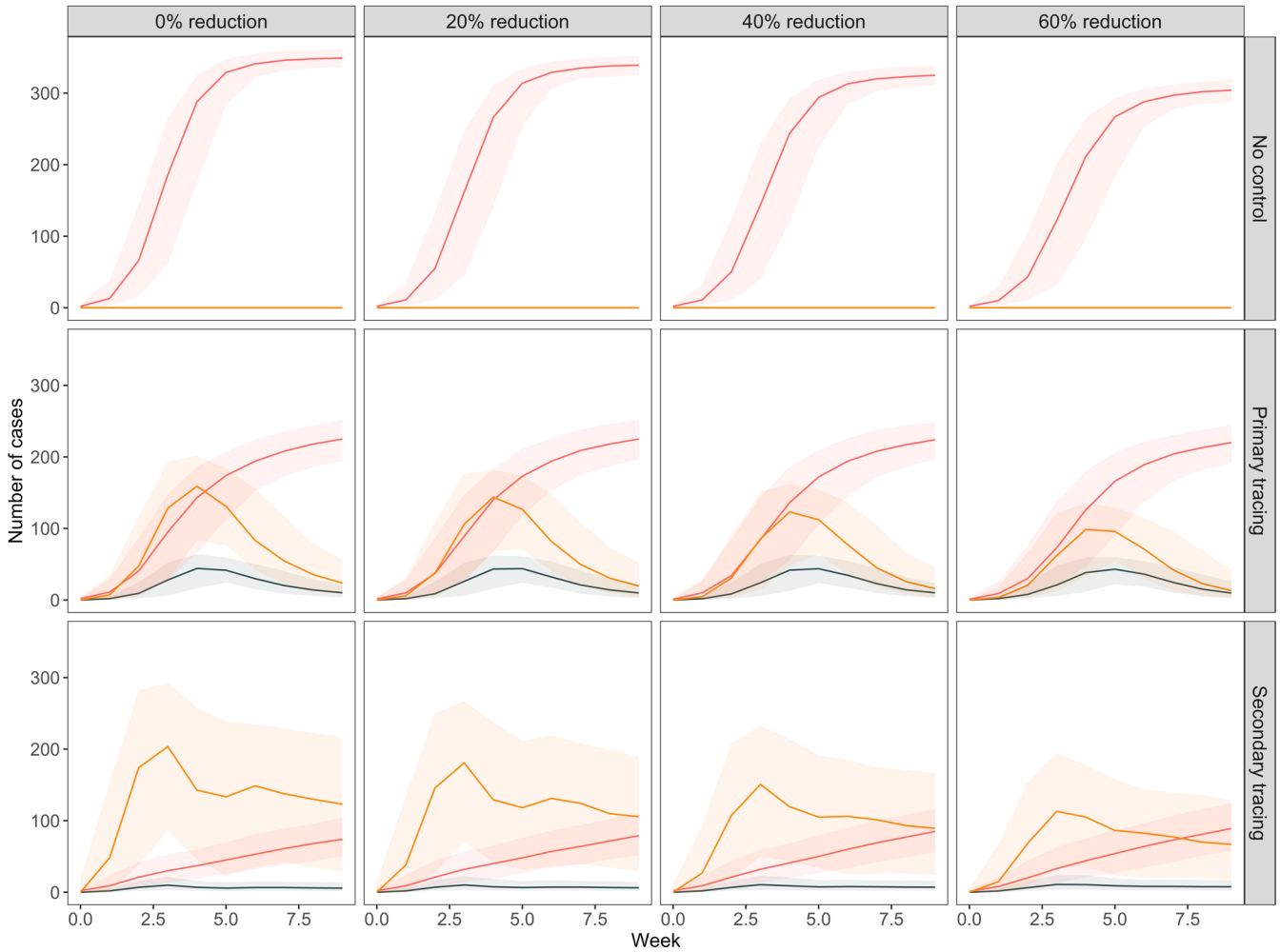
Extended Data Fig. 5 | Outbreak dynamics, isolation delay and number of initial cases. Epidemic model predictions of outbreak size and number of people isolated/quarantined in relation to the delay between case onset/tracing and isolation/quarantine (see methods for details), and the number of initial cases. Lines and shaded areas represent median and 5th-95th percentiles from 1000 simulations.



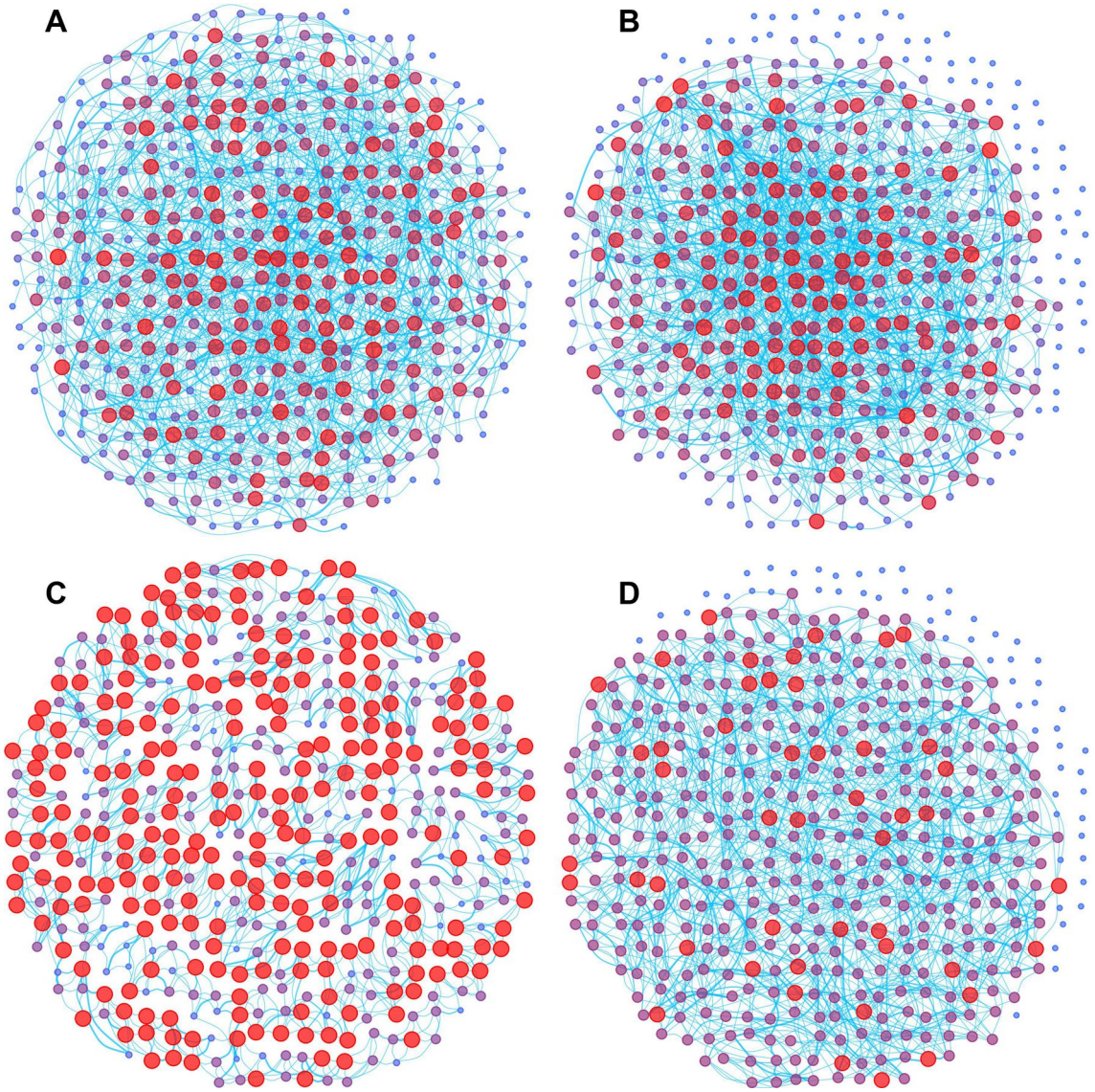
Extended Data Fig. 6 | Outbreak dynamics and outside infection rate. Epidemic model predictions of outbreak size and number of people isolated/quarantined in relation to the outside infection rate into the Haslemere network. Outside infection rate is the probability that an individual is randomly infected on a given day (see methods for details). Lines and shaded areas represent median and 5th-95th percentiles from 1000 simulations.



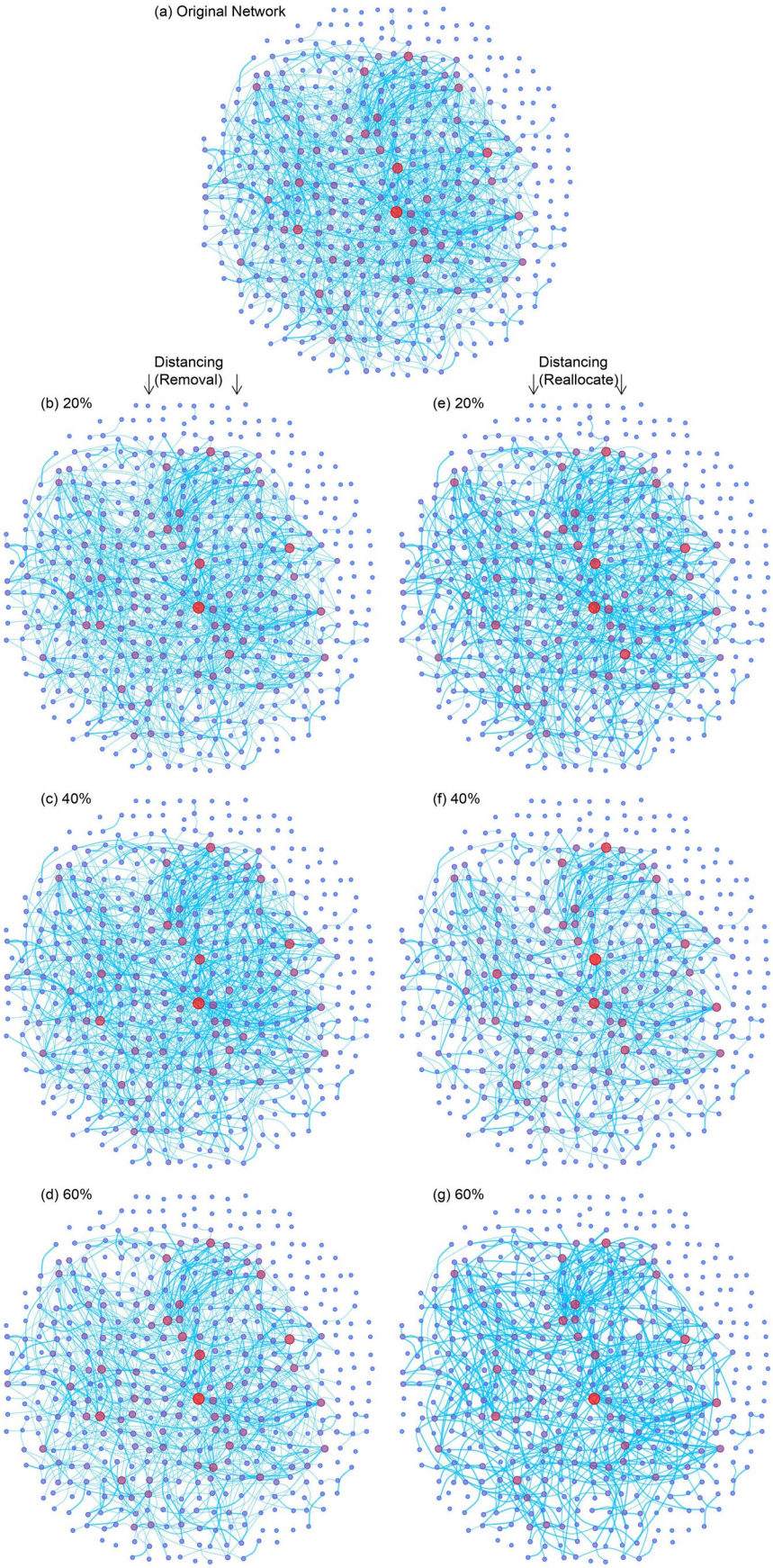
Extended Data Fig. 7 | Outbreak dynamics and dense networks. Epidemic model predictions of outbreak size and number of people isolated/quarantined under different non-pharmaceutical intervention scenarios in the Haslemere network, with distance thresholds for contacts set at **a** 7 metres, and **b** 16 metres, whereby an increased threshold results in a more densely connected network. Colours represent the 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 5th-95th percentiles from 1000 simulations.



Extended Data Fig. 8 | Outbreak dynamics and physical distancing using reassignment. Epidemic model predictions of outbreak size and number of people isolated, quarantined and tested under different levels of physical distancing in the Haslemere network, with physical distancing simulated using a reassignment method. The percentage reduction refers to the number of 'weak links' removed and reassigned (see methods and Extended Data Fig. 10). Lines and shaded areas represent median and 5th–95th percentiles from 1000 simulations.



Extended Data Fig. 9 | Null network simulations. Graphical depiction of examples of the networks of social contact generated under the four different empirically parameterised null models. Number of unique contact partners is denoted by size and colour of nodes (large & red = central) with this standardised within each panel (max node size = 3x min node size). Each panel contains the same number of unique edges in total, and the strength of the edges (denoted by thickness) is also maintained. The panel order follows that of main text Fig. 3, with **A** edge null, **b** degree null, **c** lattice null and **d** cluster null (see methods for details). Each network is organised in a fitted spring layout which is then rescaled into an equally-spaced filled circular format.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Physical distancing simulations. Graphical depiction of examples of social networks under the two different physical distancing criteria. Number of unique contact partners is denoted by size and colour of nodes (large & red = central) with this standardised within each panel (max node size = 3x min node size). Each panel contains the same nodes, and same social connections between those observed together on more than one day, but the number of unique edges occurring between individuals observed on just one day is reduced by the specified percentage. The panels show (**a**) the observed social network, (**b**–**d**) physical distancing scenario 1 at 20%, 40% and 60%, where edges that occurred on just one day are removed, and (**e**–**g**) physical distancing scenario 2 at 20%, 40% and 60% where edges that occurred on just one day are reallocated to their other connections (see methods for details). Each network is in the same network format across panels, which is a spring layout of the observed network then rescaled into an equally spaced filled circular format.

nature research

Corresponding author(s): Lewis Spurgin

Last updated by author(s): 2020/07/16

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	This study used the raw data previously published in Kissler et al.3 and are available to download at: https://github.com/skissler/haslemere. The summarized network data used here are publicly available with the code.
Data analysis	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/

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

This study used the raw data previously published in Kissler et al.3 and are available to download at: https://github.com/skissler/haslemere. The summarized network data used here are publicly available with the code.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	468
Data exclusions	*Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Replication	This study uses epidemic modelling within the context of real world social networks. Each epidemic model is carried out 1000 times. Sensitivity analysis of the model parameters is also carried out.
Randomization	N/A (not an experiment)
Blinding	N/A (not an experiment)

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	This study used data previously published in Kissler et al 2018 (as stated above) and the characteristics of the individuals were not provided as all individual identify information was removed.
Recruitment	This study used data previously published in Kissler et al 2018 (as stated above) from a population based in Haslemere, Surrey, UK. The previously published data description states that an initial 1272 individuals within Haslemere that downloaded the app (i.e. were recruited), and then 468 individuals had sufficient data points at a resolution of 1m over three full days within the focal area for further analysis. It is these 468 individuals that were provided for this particular study.
Ethics oversight	Information was provided and consent obtained from all participants in the study before the app recorded any data. The study was approved by London School of Hygiene & Tropical Medicine Observational Research Ethics Committee (ref 14400).

Note that full information on the approval of the study protocol must also be provided in the manuscript.