POLITECNICO DI TORINO
Repository ISTITUZIONALE

Effectiveness of probabilistic contact tracing in epidemic containment: The role of superspreaders and transmission path reconstruction

Original
Effectiveness of probabilistic contact tracing in epidemic containment: The role of superspreaders and transmission path reconstruction / Muntoni, A.P., Mazza, F., Braunstein, A., Catania, G., Dall'Asta, L.. - In: PNAS NEXUS. - ISSN 2752-6542. - 3:9(2024). [10.1093/pnasnexus/pgae377]

Availability:
This version is available at: 11583/2993107 since: 2024-10-07T07:26:52Z

Publisher:
Oxford University Press - National Academy of Sciences

Published
DOI:10.1093/pnasnexus/pgae377

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

26 November 2026

PNAS Nexus, 2024, 3, pgae377
https://doi.org/10.1093/pnasnexus/pgae377
Advance access publication 3 September 2024
Research Report

Effectiveness of probabilistic contact tracing in epidemic containment: The role of superspreaders and transmission path reconstruction

Anna Paola Muntoni[a,b,*], Fabio Mazza[a,c], Alfredo Braunstein[a,b], Giovanni Catania[d] and Luca Dall'Asta[a,b,e]

[a]Department of Applied Science and Technology, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy
[b]Statistical inference and computational biology, Italian Institute for Genomic Medicine, c/o IRCSS, Candiolo 10060, Italy
[c]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Via Ponzio 34/5, Milano 20133, Italy
[d]Departamento de Física Teórica, Universidad Complutense, Madrid 28040, Spain
[e]Collegio Carlo Alberto, P.za Arbarello 8, Torino 10122, Italy
*To whom correspondence should be addressed: Email: anna.muntoni@polito.it
Edited By Sandro Galea

Abstract

The recent COVID-19 pandemic underscores the significance of early stage nonpharmacological intervention strategies. The widespread use of masks and the systematic implementation of contact tracing strategies provide a potentially equally effective and socially less impactful alternative to more conventional approaches, such as large-scale mobility restrictions. However, manual contact tracing faces strong limitations in accessing the network of contacts, and the scalability of currently implemented protocols for smartphone-based digital contact tracing becomes impractical during the rapid expansion phases of the outbreaks, due to the surge in exposure notifications and associated tests. A substantial improvement in digital contact tracing can be obtained through the integration of probabilistic techniques for risk assessment that can more effectively guide the allocation of diagnostic tests. In this study, we first quantitatively analyze the diagnostic and social costs associated with these containment measures based on contact tracing, employing three state-of-the-art models of SARS-CoV-2 spreading. Our results suggest that probabilistic techniques allow for more effective mitigation at a lower cost. Secondly, our findings reveal a remarkable efficacy of probabilistic contact-tracing techniques in performing backward and multistep tracing and capturing superspreading events.

Keywords: contact tracing, message passing, epidemic containment, superspreaders, statistical inference

Significance Statement

The recent experience of the COVID-19 pandemic, especially at the early stage of propagation, has highlighted the importance of non-pharmaceutical interventions to help contain or mitigate epidemic outbreaks while maintaining a low social and economic impact. A promising intervention of this type is a targeted test-isolation protocol guided by digital contact tracing (DCT). However, standard implementation of DCT rapidly becomes impractical in a rapid spreading phase where the number of notifications and the number of associated recommended tests grows dramatically. Probabilistic contact tracing is an alternative to standard protocols that is more robust and allows for higher mitigation effectiveness at lower social cost and thriftier usage of diagnostic resources. The success of these techniques relies on the effective detection of superspreading events and a correct reconstruction of transmission paths.

Introduction

The recent experience of the COVID-19 pandemic has shown that mobility restrictions and lockdowns can have severe social and economic consequences (1). In light of the potential unavailability of vaccines, particularly in the early stages of a pandemic, it is then imperative to develop and implement nonpharmacological intervention measures capable of ensuring the containment or gradual slowing down of epidemic outbreaks while concurrently preserving economic and social activities (2, 3). Together with increased attention to hygiene and the use of masks, contact tracing represents the most promising nonpharmacological measure for this purpose (4), and has been successfully employed to identify and eradicate small outbreaks of COVID-19 (5, 6). Manual contact tracing (MCT) becomes impractical for large epidemic outbreaks, implying high costs and temporal delays (4, 7, 8). Moreover, MCT is unlikely to discover contacts outside of immediate family or close relationships (9, 10). Building on previous studies related to the Ebola virus disease (11, 12), it has been

Competing Interest: The authors declare no competing interest.
Received: July 11, 2024. **Accepted:** August 15, 2024
© The Author(s) 2024. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Downloaded from https://academic.oup.com/pnasnexus/article/3/9/pgae377/7748389 by guest on 07 October 2024

argued that such limitations could be overcome with the systematic use of automated contact tracing procedures, which could scale up to the case of large outbreaks and favor the discovery of potentially infectious contacts even among occasional ones (13, 14) (see also (15)). Indeed, aggressive containment policies based on digital contact tracing (DCT) technologies, such as smartphone apps and GPS beacons, proved effective during the first wave of COVID-19 in countries like Taiwan (16), South Korea (17), China (18), and Singapore (19). These techniques sparked debates in Western countries on the threat of individual privacy (20–22) and the need for voluntary adoption of contact-tracing apps by a large portion of the population (13, 23, 24). Privacy-preserving protocols for digital contact tracing (DCT) have been introduced, using either centralized (25–27) or distributed (28–30) approaches, primarily relying on Bluetooth low-energy (BLE) communication to detect physical proximity without geolocation. The analysis of data obtained from early implementations of DCT apps indicates a tangible contribution to epidemic containment, providing an additional quantitative and qualitative advantage over MCT (31–34).

In most DCT apps, exposure notifications are triggered for every contact with individuals who have tested positive, irrespective of a variety of factors determining the risk associated with the contact. As a consequence, the proliferation of exposure notifications and quarantines, responsible for the reduction in the number of infected individuals, can lead to very high social costs (e.g. the number of isolated individuals) and economic costs (e.g. the number of diagnostic tests used) (35–37). A crucial step towards improving the efficacy of DCT and reducing notification redundancy is represented by probabilistic contact tracing methods, which could naturally account for multiple exposures (38–41). Using a Bayesian framework that incorporates all available data on individuals who tested positive (or negative), Baker et al. (39) proposed an efficient distributed method, based on Belief Propagation (42, 43), to compute the individual probabilities of infection. This information can be leveraged by the contact tracing app to determine individual risk levels presented to the users, favoring self-isolation and more efficient testing strategies. In the present study, following the approach of Baker et al. (39), we demonstrate the superiority of probabilistic contact tracing methods over standard ones, both in terms of higher containment capacity and lower cost-to-benefit ratio. This is done through a comparative analysis using different epidemic simulators (44–46), obtaining results that are robust across various disease transmission models and parameter ranges.

This study also offers the opportunity to delve deeper into the mechanisms and causal relationships that control automated contact tracing, investigating the reasons behind the claimed superiority of probabilistic methods. As positive tested individuals are more likely to come from contagion clusters than to generate them (47), it is believed that the detection of sources of individual infections (backward tracing) and superspreading events can significantly improve containment strategies, especially in the presence of overdispersed secondary infections (48, 49), a common trait of modern diseases such as COVID-19 (50–55) and mpox (56–58). Countries like Japan (59), South Korea (60), and Uruguay (61) are credited with successfully implementing backward tracing in their contact tracing campaigns. However, current app-based DCT implementations predominantly engage in simple forward tracing (62), where tracked individuals are primarily those who could have been exposed to someone who has tested positive. Innovative DCT methods based on statistical inference (39), which ground their predictive power on reconstructing causal relationships in transmission paths (42), are instead expected to more efficiently discover multistep forward and backward traces and capture superspreading events. This is here quantitatively demonstrated by analyzing these features for various contact tracing strategies across different epidemic models in the early-containment phase, providing a possible explanation of the superior containment ability of probabilistic contact tracing.

Results

Mathematical models of epidemic spreading are largely used to forecast the evolution of outbreaks at different spatial and temporal scales, to evaluate the effects of public health interventions, and ultimately to guide governments' decisions (63–65). In this respect, agent-based models provide stylized but sufficiently reliable representations of the actual contact networks on which contagion between individuals could take place, thus becoming a natural and necessary tool for analyzing the consequences of nonpharmaceutical intervention strategies based on contact tracing. Among the abundance of agent-based models proposed during the first waves of the COVID-19 pandemic (44–46, 66–68), some of them can be considered exemplary for formulating a critical analysis of the containment capabilities of the different contact tracing methods and evaluate their cost-to-benefit ratio. The agent-based models analyzed in the present work, namely the OpenABM model by Hinch et al. (44), Covasim by Kerr et al. (45) and the Spatiotemporal Epidemic Model (StEM) by Lorch et al. (46), can be considered rather simple generalizations of the Susceptible-Exposed-Infected-Recovered (SEIR) model, in which additional states are included to account for different levels of symptomaticity and disease severity. Agent populations are endowed with realistic features, including demographic data and different layers of social interactions, also obtained from simulated mobility (see Methods and the Supplementary Material for details). As a consequence, such models are capable of reproducing the empirically observed non-Poissonian statistics and overdispersion in contact patterns and individual viral loads.

These three agent-based models, each characterized by their unique attributes, serve as an ideal platform to assess the efficacy of contact tracing methods based on statistical inference, demonstrating their superiority in comparison to conventional test-trace-quarantine approaches. The probabilistic methods under study are those appearing in Baker et al. (39), namely Simple Mean Field (SMF) and Belief Propagation (BP). For comparison, other contact tracing methods are considered: a basic form of DCT, and a more advanced "informed" contact tracing (ICT) approach that leverages all available information from medical test results. Additionally, for Covasim, we employed Test-Trace-Quarantine (TTQ), the integrated containment method presented in the work by Kerr et al. (69). This method employs information about the symptomatic status of the tested individuals; even though encoding this data into BP is always possible, we do not use this information while running BP, SMF, DCT, and ICT to allow for a fair comparison among the four methods. The Methods section provides a brief overview of the contact tracing algorithms (see Supplementary Material for further details). The containment effectiveness of these different contact tracing methods is evaluated by a quantitative study across various intervention scenarios generated using these three agent-based models. Our analysis demonstrates that contact tracing based on statistical inference techniques facilitates effective mitigation at low medical costs, measured in terms of diagnostic tests, and social costs, quantified by the fraction of the population subjected to quarantine. Finally, tracing techniques based on statistical

Downloaded from https://academic.oup.com/pnasnexus/article/3/9/pgae377/7748389 by guest on 07 October 2024

inference are shown to outperform other approaches in effectively tracing both backward and forward transmissions and therefore in identifying superspreading events associated with the overdispersion of secondary infections.

Epidemic containment

DCT-based strategies possess a remarkable capability to contain the spread of epidemics by reducing their impact. This was recently demonstrated within the realistic framework provided by OpenABM (39). A similar analysis is carried out here on several instances of epidemic spreading generated using Covasim and StEM from a small initial number of infected individuals (patient zeros). The contact tracing protocol involves daily testing of a fixed fraction of symptomatic individuals. Different contact-tracing methods exploit the initial phase to gather information and update a ranking of potentially infected individuals. Starting from the first day of intervention t_i, an additional number of individuals is tested daily according to the risk predictions provided by the different methods. Those who test positive are subsequently confined. To formulate the ranking, each contact tracing algorithm incorporates the diagnostic test results and the contacts collected by the underlying contact tracing app over a predefined period. We assume that the app gathers the same information for all contact tracing methods, contingent on the app's adoption fraction (AF) within the population (assumed to be $AF = 1.0$ here). The effects of lower adoption fractions ($AF < 1$) were investigated in (39) for the case of OpenABM, but we expect similar behaviors for the two other models studied here. Note, however, that even for $AF = 1$, the transmission network may be significantly different than the contact network detected by the app. In StEM, some exogenous transmissions are added within the simulation, and in Covasim the relative transmission among individuals, i.e. the weights of the transmission network, is highly heterogeneous and inaccessible to the contact tracing app and inference method (see Methods). The test results are subject to error due to a non-zero false-negative rate (f_N). In our simulations, we set $f_N = 0.285$, an estimated value derived from published data (70), representing a relatively high false-negative rate associated with rapid COVID-19 tests that provide quick and affordable, but less accurate contagion assessment.

A standard testing strategy, applicable to all contact tracing methods, entails performing a fixed number of tests per day. In the strategies labeled as DCT, ICT, SMF, and BP $\tau = 7$ the number of tests is fixed to $N_{test} = 220$ for StEM and $N_{test} = 230$ for Covasim. However, it is worth noting that probabilistic contact tracing methods like BP and SMF allow for an alternative testing strategy. This approach involves observing individuals whose estimated probability of being infected exceeds a threshold value (p_{th}). In this case, the number of tests based on the ranking changes adaptively over time. For StEM, we set $p_{th} = 1 \times 10^{-4}$, and $p_{th} = 5 \times 10^{-5}$ for BP and SMF respectively, while for Covasim, we set $p_{th} = 3 \times 10^{-3}$ for SMF and $p_{th} = 5 \times 10^{-3}$ for BP (see Figures S4 and S5 for the performances of the two algorithms under varying thresholds). The significant advantage inherent in this testing strategy is that each test is performed based on an estimate of the individual's medical status. This has a twofold impact. First, when no individual is eligible for testing, no diagnostic test is administered, leading to a more parsimonious use of medical resources compared to the fixed N_{test} setting. Second, this approach addresses ethical considerations by encouraging testing only for individuals with a high likelihood of being infected.

To quantify the effectiveness of each containment policy and to set the stage for the analysis carried out in the next sections, Figure 1 shows the effective reproduction number R_t (refer to

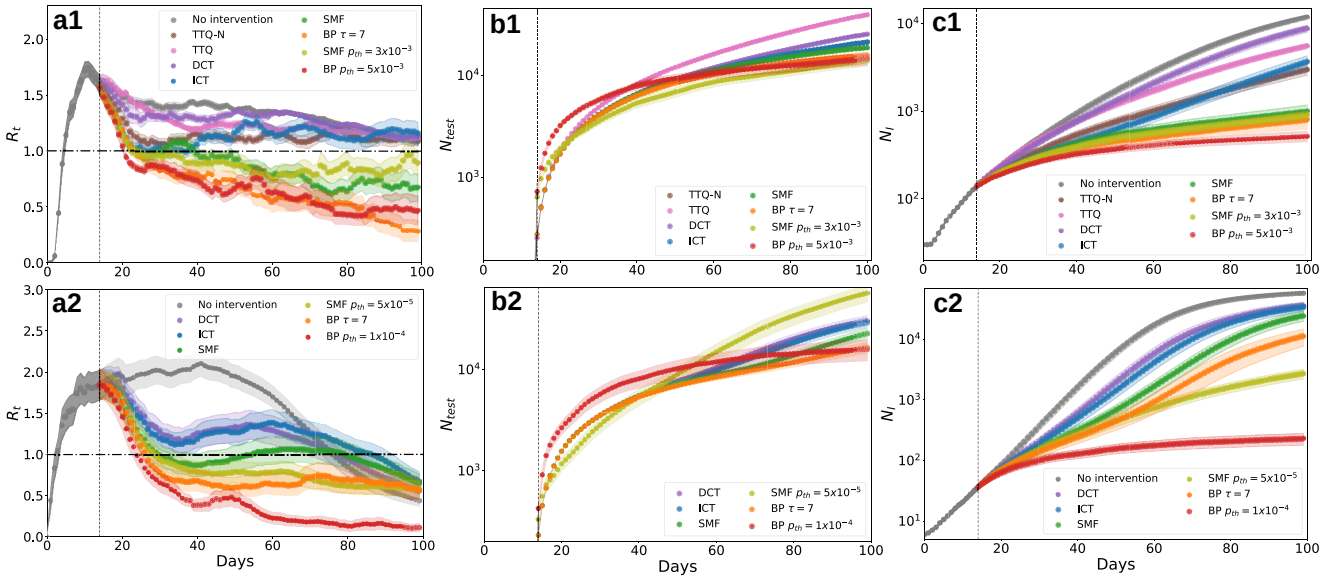

Fig. 1. Effective epidemic mitigation. Columns labeled a), b), and c) show, respectively, the behavior in time of the effective reproduction number R_t (see Supplementary Material for a detailed description), the cumulative number of diagnostic tests, and the cumulative number of infected individuals. For the Covasim model (first row), simulations are run on a population of 70,000 people, for $T = 100$ days. Each simulation starts with $N_{pz} = 30$ patients zero, all in the exposed state, and each day half of the unidentified symptomatic individuals are observed ($p_{sym} = 50\%$), while tracing-based interventions start after $t_i = 14$ days. For the StEM model (second row), simulations are performed on the urban area of Tübingen for $T = 100$ days, and the number of initial cases N_{pz} is 6 (1 in the exposed state, 2 in the asymptomatic state, and 3 of them are presymptomatic individuals). The same fraction of the symptomatic individuals is observed ($p_{sym} = 50\%$), with interventions starting at $t_i = 14$. In the StEM model, households are confined whenever a member is tested positive. Lines reflect the average behavior of the metrics computed from 20 realizations of the Covasim population model and 30 realizations of the StEM mobility model. The shaded regions indicate the associated standard error.

Downloaded from https://academic.oup.com/pnasnexus/article/3/9/pgae377/7748389 by guest on 07 October 2024

the Methods section for a definition), the cumulative number N_I of the infected individuals and the cumulative number N_{test} of performed tests (included those administered to symptomatic individuals) over time. In both models, all nonprobabilistic methods face challenges in sustaining R_t below one, even in the long run, whereas BP, and to a lesser extent SMF, prove to be more adept at achieving this goal swiftly.

Cost–benefit analysis

In addition to the economic costs associated with medical tests, nonpharmacological epidemic containment policies also impose a social cost due to mobility restrictions. This cost can be quantified by measuring the cumulative number, or percentage, N_Q of individuals in quarantine as a result of different contact tracing strategies. This quantity is then compared to the effective reduction in epidemic spread, defined as one minus the ratio between the infected individuals in a mitigated scenario and that in an uncontrolled regime, where only a fixed percentage of symptomatic individuals are tested and quarantined. The values of reduction are computed when the number of infected individuals in uncontrolled simulations reaches a plateau (which happens roughly at $T = 100$ for the StEM and at $T = 150$ for Covasim). Higher values of reduction indicate better containment performance. This cost-to-benefit analysis was first introduced in Ref. (35), where the authors investigated a theoretical expectation of the number of required quarantines to achieve a specific reduction in the final epidemic size when manual and DCT is applied. For the comparison, the settings described in Figure 1 for both the Covasim model and StEM are adopted. Figure 2(a.1) and (b.1) shows the reduction measure defined above as a function of the number of tests performed daily, for Covasim and StEM, respectively. The size of the markers reflects the cumulative number of quarantined individuals resulting from the employed contact tracing strategy (larger dots correspond to larger numbers). The color gradient represents the number of daily tests conducted during the simulation, with darker colors indicating a larger number of observations. As clearly shown by these results, the two probabilistic methods (i.e. SMF and BP) always reach higher performances in terms of reduction at a fixed number of medical tests.

Similarly, panels (a.2) and (b.2) display the percentage of individuals in quarantine generated by the intervention strategy (excluding isolation associated with symptomatic individuals) as a function of the reduction (see Figure S6 for the plot of the number of confined individuals as a function of the number of daily tests). Regardless of the number of available rapid tests, the number of confined individuals is significantly smaller for probabilistic contact tracing techniques (BP and SMF) than for the others (DCT and ICT). This suggests that not only the two techniques are preferable in terms of effectiveness, but they also incur a lower social cost as fewer individuals need to be isolated. Our numerical estimates appear qualitatively similar to the results in Ref. (35) where the authors predicted a behavior similar to a downward opening parabola for the number of quarantines as a function of the reduction. In our case, we stress that BP-based curves are always associated with lower values of the isolated cases N_Q at fixed reduction values. The color gradient in panels (a.2), and (b.2) also reveals that this result is achieved at a lower diagnostic cost as the number of necessary tests to reach the same performance in terms of reduction, is lower than that used by the other methods. This behavior is particularly pronounced in StEM: BP obtains a reduction greater than 0.8 using about 400 daily tests while SMF needs at

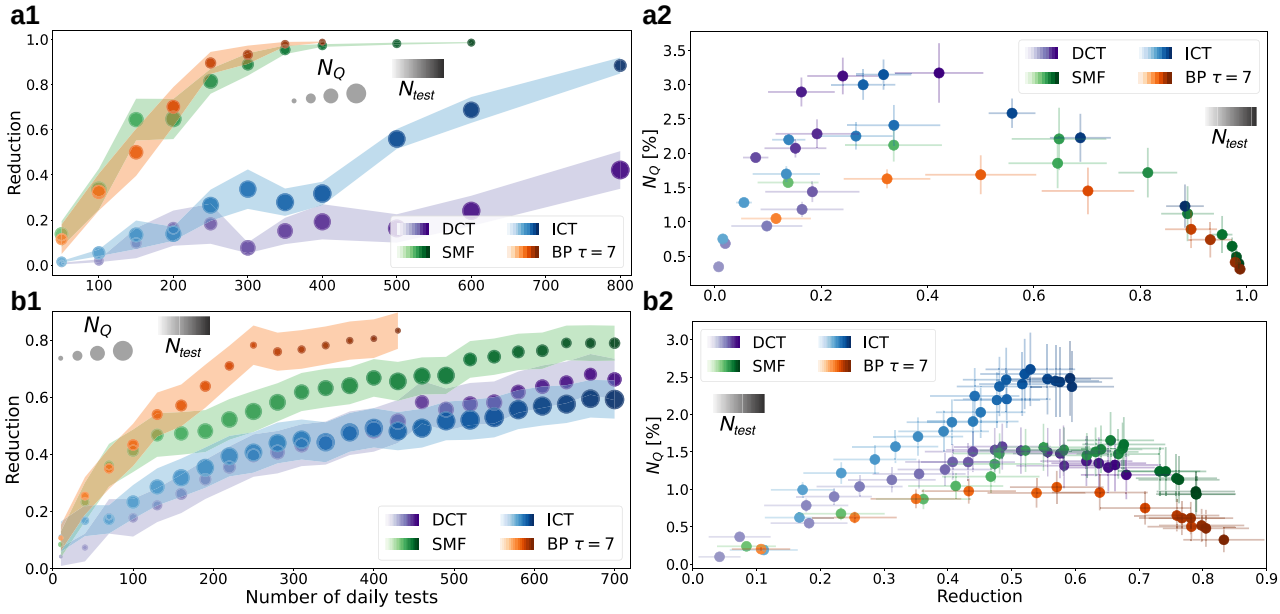

Fig. 2. Spreading reduction, social, and diagnostic cost. Panels (a.1) and (b.1) show the reduction measure of the epidemic spreading as a function of the number of medical tests performed daily during the simulations; panels (a.2) and (b.2) display N_Q, the percentage of the confined individuals due to the different confinement strategies as a function of the reduction (see Ref. (35)). These quantities are computed for $T = 100$ and $T = 150$ for StEM and Covasim respectively, when the number of infected individuals reaches a plateau in the corresponding uncontrolled simulations. For StEM (Covasim), the population has a size of 90, 546 (70,000) individuals (see Methods). The panels on the top display the two measures associated with the Covasim model, while the panels on the bottom show the results while running StEM dynamics. The reduction measure is formally defined as the difference between the cumulative number of infected in an unconstrained propagation (where only the fixed percentage of symptomatic is confined) and the mitigated one, normalized by the cumulative number of infected in unconstrained dynamics. The higher the reduction, the more effective the containment measure. The size of the markers in panels (a.1) and (b.1) is proportional to the number of quarantines (the quantity plotted in the y-axis of the (a.2) and (b.2) panels), the larger the dots, the larger the number of confined individuals. The color code used in all the panels mirrors the number of tests performed on a daily basis: the darker the color, the larger this number.

Downloaded from https://academic.oup.com/pnasnexus/article/3/9/pgae377/7748389 by guest on 07 October 2024

least 700 observations, and DCT and ICT never reach this value with the number of tests considered for this experiment (see panel (b.1)).

Overdispersion and superspreaders

Probabilistic-based tracing methods also exhibit a remarkable ability to effectively detect superspreaders. Superspreading transmission can have distinct origins, contingent on the properties of both the viral disease and the underlying population. This diversity is represented and exemplified by the three agent-based models under study. In OpenABM (13), superspreading events occur due to an overdispersed distribution of contacts in one of the three network layers used to model the population structure. Similarly, in StEM (46), overdispersion arises naturally from the contact graph, as a result of realistic mobility simulations based on geolocalized data within an urban area. In both cases, the empirical distribution of the number of infections exhibits significant non-Poissonian statistics, characterized by a variance-to-mean ratio (VMR) larger than one (refer to the Supplementary Material for further details). For these two models, individuals who infect at least seven contacts within their infectious time window are identified as superspreaders, following the definition provided in Wong et al. (71). In contrast, in Covasim (45), the overdispersion of infections directly arises from the properties of individual viral load, which is drawn from a fat-tailed distribution (see Supplementary Material): superspreaders can therefore be identified by looking at the individual relative transmission intensity T_{rel}, a quenched parameter not accessible to the tracing methods. In particular, in each simulation, individuals displaying $T_{rel} \geq 5$ are classified as superspreaders.

The ability of the different contact tracing methods to detect superspreaders among the infected individuals is evaluated through numerical experiment employing the following procedure: in each epidemic realization, the propagation is allowed to evolve freely without intervention up to a time T, whereupon the contact tracing methods are applied once, and the corresponding ranking of potentially infected individuals is collected. The value of T is here chosen to be of the order of a few weeks, representing the typical time window for which contact information can be retained in DCT applications (39). To mimic a realistic setting, we assume that individuals showing symptoms spontaneously take tests and their results are collected by the contact tracing app. This is encoded in our simulations by observing a fixed fraction of the symptomatic individuals daily (see caption of Figure 3 for additional details). Individuals identified by means of the different contact tracing methods, and ranked based on their epidemic risk, are then classified according to their true infection status, obtaining corresponding ROC curves. To specifically study the detection of superspreaders (and not other infected individuals), only the subset consisting of (a posteriori determined and nonobserved) superspreaders and susceptible individuals at time T was considered (refer to Figure 3a for a schematic representation of the setup). Superspreaders who recovered before time T were not taken into account, as their number is negligible after

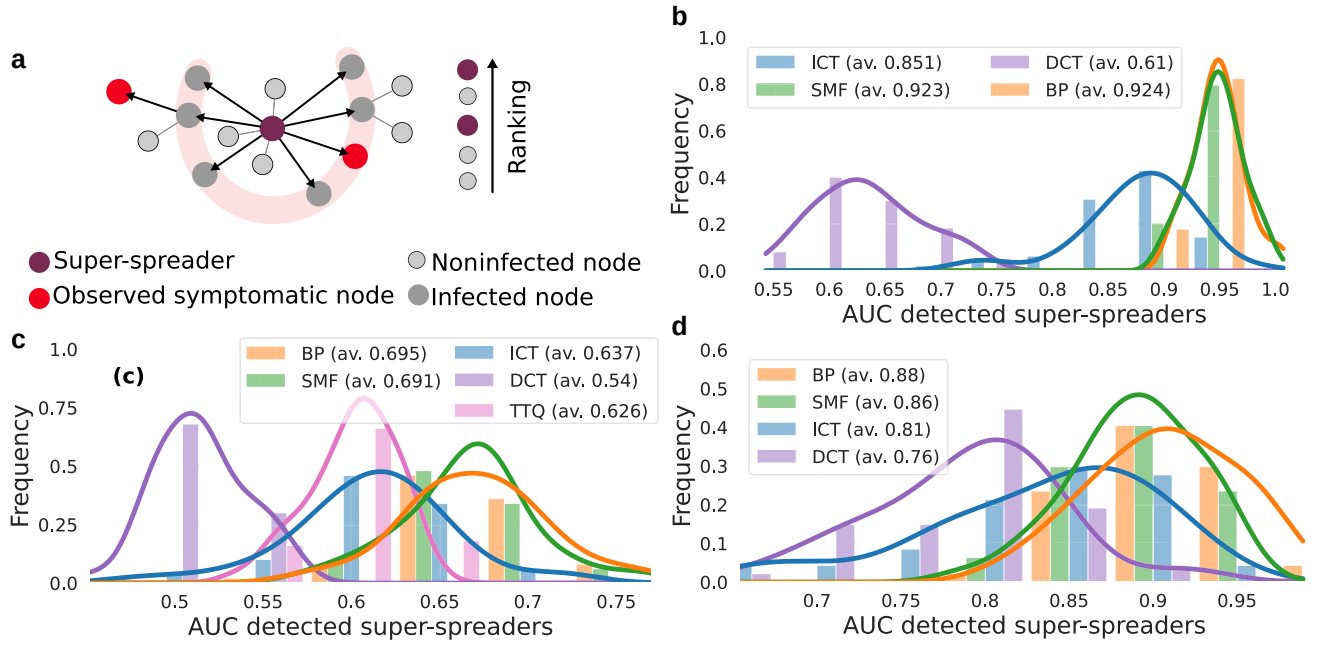

Fig. 3. Detection of the superspreaders. a) Schematic representation of the experimental setup. A posteriori, the superspreader individuals (the purple nodes) are identified as those responsible for over-dispersed transmissions (see the main text for a proper definition for the three models), here marked as the nodes within the pink shadow. To fairly evaluate the ability of each contact tracing method to detect superspreaders the ROC curves are built only for a subset of the individuals composed of the true superspreaders and susceptible individuals (small light gray nodes). Information about the epidemic dynamics entirely comes from the contact network and the daily observation of a fixed fraction of symptomatic (red nodes). The methods employed to compute the ROC curves are Belief Propagation (BP), Simple Mean Field (SMF), Informed Contact Tracing (ICT), DCT, Trace-Test-Quarantine (TTQ). The statistics of the AUC associated with the ROC curves obtained by different methods are shown for b) OpenABM, c) Covasim, and d) StEM. Lines are kernel density estimation plots used as guides for the eyes, while mean AUC values are reported in the legend. All parameters used in these simulations are the same as used in the epidemic containment results, except for the time T, the number of patients zero N_{pz}, and the probability of self-testing. These numbers have been tuned to ensure that the maximum number of true positives in the ROC curves is at least a few tens. In particular, the duration of the free epidemic propagation before estimation is set to $T = 15$ for StEM, $T = 30$ for Covasim, and $T = 20$ for OpenABM. The number of initially infected individuals is set to $N_{pz} = 200$ for StEM, $N_{pz} = 90$ for Covasim and $N_{pz} = 100$ for OpenABM. The fraction of observed symptomatic individuals is set to $p_{sym} = 0.1$ for StEM and for Covasim, while for OpenABM all severe symptomatic individuals are observed ($p_{ssym} = 1.0$) together with a fraction $p_{msym} = 0.3$ of mild ones.

Downloaded from https://academic.oup.com/pnasnexus/article/3/9/pgae377/7748389 by guest on 07 October 2024

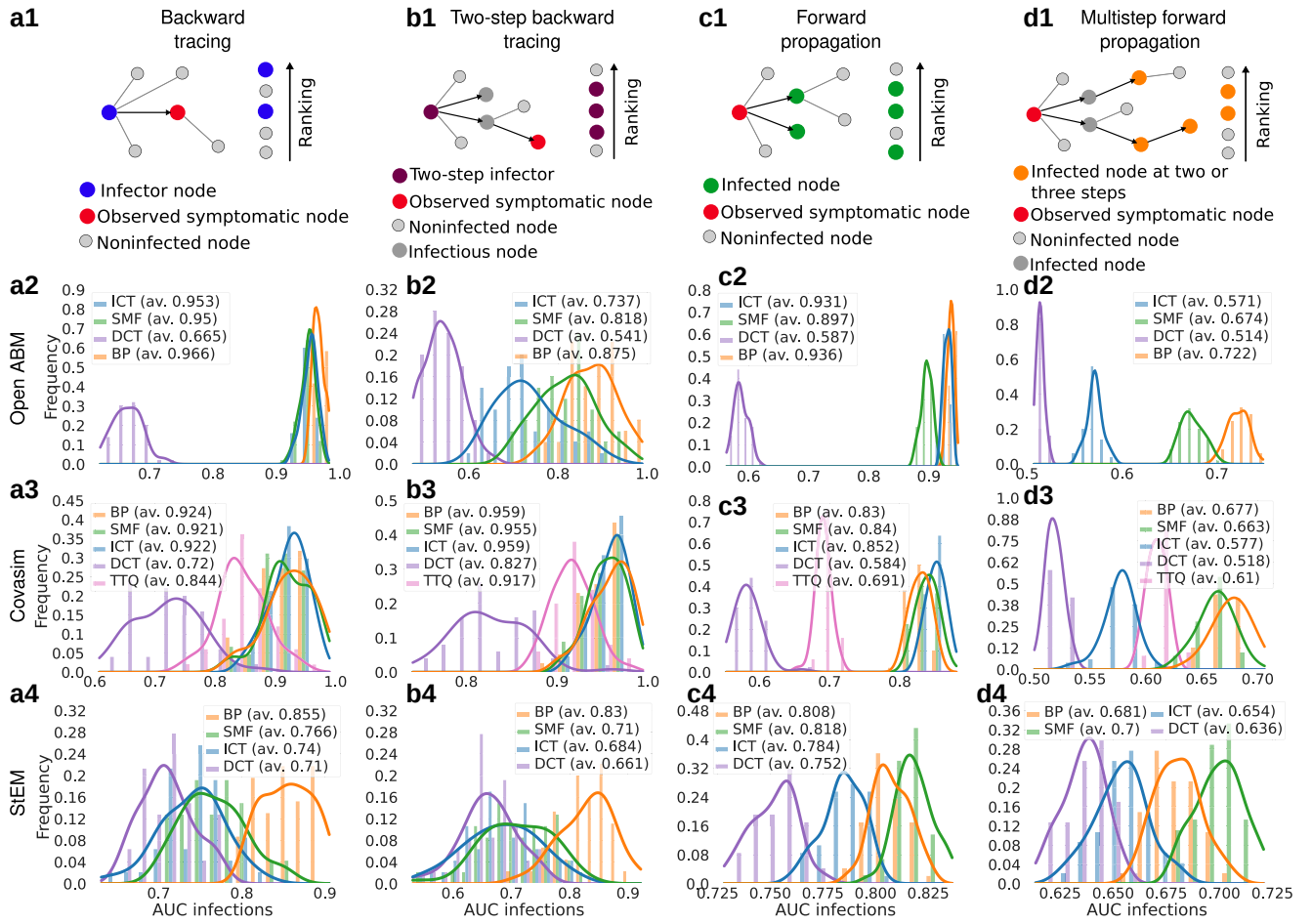

Fig. 4. Detection of the one-step, two-step backward, and one-step, multistep forward tracing. Panels (a.1), (b.1), (c.1), and (d.1) show a schematic representation of the one-step, two-step backward, one-step, and multistep forward transmissions respectively. See the main text for a formal definition. The second, third, and fourth rows show the histogram of the AUC associated with the detection of the four types of infected individuals, for OpenABM, Covasim, and StEM, respectively. The methods used to obtain the ROC curves are Belief Propagation (BP), Simple Mean Field (SMF), Informed Contact Tracing (ICT), DCT, and Trace-Test-Quarantine (TTQ). The simulation set-up used for these results is the same exploited for the detection of the superspreaders illustrated in Figure 3. The average AUC is reported in the legend for the methods, while the lines report kernel density estimates to guide the visualization of the histograms.

T days. Figure 3b–d illustrates the empirical distributions of the area under the curve (AUC) obtained from different contact tracing methods across multiple epidemic realizations for OpenABM, Covasim, and StEM. In all three models, probabilistic methods (SMF and BP) turn out to better differentiate between noninfected and superspreaders, as indicated by both the distribution of the AUC (it is significantly shifted towards larger values for SMF and BP) and the average value of the AUC shown in the legend. Conversely, the distributions associated with ICT, DCT (and TTQ for Covasim) predictions are concentrated at lower values, confirming that nonprobabilistic algorithms are less effective in tracing superspreader exposures.

Backward and forward tracing

One of the inherent difficulties in contact tracing is determining the direction of infection among confirmed cases. While tracing new infections (forward tracing) is relatively easier, a more complex task is to trace the source of the observed infections (backward tracing). The ability to identify transmissions backward is crucial for detecting superspreaders and effectively mitigating the spread of an outbreak (48, 49). To further emphasize the advantages of probabilistic contact tracing methods like SMF and BP, it is valuable to assess their ability to identify secondary and tertiary infections, i.e. new infections that occur two or three steps away from the observed individuals in the transmission history. The experimental setup employed in Figure 4 consists of the following: for each epidemic realization, the propagation is allowed to evolve without intervention until a time T, and a small fraction of symptomatic individuals is observed daily. The backward propagators are defined as the sources of infection for the observed symptomatic individuals (depicted as blue dots in Figure 4a.1); their infectors instead identify the two-step backward propagators (see Figure 4b.1). Forward propagators are defined as the secondary infections of observed individuals (represented by green nodes in Figure 4c.1). New infections occurring at two and three steps from the observed individuals are shown as orange nodes in the example presented in Figure 4d.1. To quantify the performances of the ranking methods, a comparison is made using the AUC associated with the classification of the infected individuals in a restricted set, where the false positive set comprises all noninfected individuals (light gray nodes in Figure 4a.1–d.1) while the true positive set consists of the unobserved one-step and two-step backward infectors, forward infections, or new infections occurring at steps two and three, respectively. Other infected individuals not belonging to these three

Downloaded from https://academic.oup.com/pnasnexus/article/3/9/pgae377/7748389 by guest on 07 October 2024

categories (e.g. tested-positive individuals, represented by red nodes in Figure 4a.1–4d.1) are not considered. Although the performances vary across the three epidemic models, the results in Figure 4 demonstrate that probabilistic models such as BP and SMF are highly effective in identifying transmissions forward and backward. For OpenABM (panels a.2–d.2) and StEM (panels a.4–d.4) probabilistic contact tracing methods outperform the others, particularly when detecting one-step, two-step backward. In the case of Covasim (panels a.3–d.3), probabilistic methods appear to play a crucial role mainly in detecting multistep forward transmissions, while their performances are similar to ICT in detecting backward and one-step forward transmissions. In these last three scenarios, simpler and less computationally expensive nonprobabilistic contact tracing methods (DCT and TTQ) do not reach the same AUC values achieved by ICT. We stress that although TTQ includes additional information about the symptomatic status of the individuals, it still does not attain the accuracy of probabilistic methods.

Discussion

Contact tracing stands out as a compelling strategy to support and improve the effectiveness of common nonpharmaceutical mitigation measures, such as social distancing, the use of masks, and other hygiene practices, in order to contain the spread of emerging viral diseases. This approach has the potential to prevent the need for measures with significant socioeconomic impacts, such as lockdowns. In particular, DCT overcomes the limitation of manual contact tracing by encompassing the ability to detect presymptomatic and asymptomatic individuals outside of close and known relationships with tested individuals, a key aspect in the prevention of highly contagious diseases, such as COVID-19. The primary drawback of current implementations of digital contact tracing is that the volume of exposure notifications delivered drastically grows with the outbreak size. Consequently, the number of individuals flagged for testing grows substantially, rendering the overall procedure impractical. A potential solution to this challenge involves enhancing individual-based epidemic risk assessment and using it to guide selective test-trace-isolate/quarantine strategies. This can be accomplished by integrating contact tracing with distributed statistical inference methods, capable of reconstructing contagion channels from locally collected information and providing a more accurate estimate of individual risk (39). These algorithms can be implemented in a privacy-preserving distributed way through smartphone apps based on current technology and without the need for centralized calculations. It was estimated that when implementing BP or SMF the amount of information sent and received between two users could be approximately 1 megabyte (MB) or 2 kilobytes (KB) per day, respectively (39).

The present work builds on this direction providing a quantitative comparative analysis of the performance of different contact tracing methods across various epidemic regimes using three distinct epidemic models recently developed for COVID-19. In all scenarios under study, probabilistic contact-tracing methods effectively curb ongoing outbreaks, as indicated by the rapid reduction of the effective reproduction number below the critical value of one. This is achieved with a substantially lower cumulative number of infected individuals compared to other methods, all while incurring a similar or significantly reduced deployment of testing resources. The cost (number of quarantines) versus benefit (outbreak reduction) analysis clearly shows a more favorable ratio for probabilistic contact-tracing methods, in particular for BP. Note that experiments conducted in this work utilize imperfect information about the underlying contact network (specifically, not all exposure events are assumed in StEM to be traceable, and contact strength is highly variable in Covasim but this information is not available to the inference algorithms). Other regimes with more uncertainty on the contact network will be investigated in future developments.

The numerical experiments also revealed that probabilistic methods are better suited than others to detect superspreading events, whether stemming from an innate variety of transmissibility or the heterogeneity of the contact network. This capability is crucial for the containment of emerging viral diseases characterized by overdispersion in secondary infections. Due to the presence of superspreaders, it becomes essential to work backward to identify the sources of infection for observed cases, as many individuals are likely infected by someone who also transmitted the virus to other people. In this respect, probabilistic contact tracing methods were found to outperform other methods in correctly reconstructing infection channels by one-step and multistep backward and forward tracing. On one hand, these results provide valuable insights into the mechanisms and causal patterns that govern the detailed functioning of contact tracing. On the other hand, the emerged superiority of probabilistic methods demonstrates that greater effectiveness in detecting superspreading events and backward and multistep causal relationships is crucial for successful epidemic containment strategies.

Our findings also have several practical implications. As pointed out by the threshold-based probabilistic methods, early interventions using a large number of tests appear to be always advantageous. This approach ensures a better assessment of the population-wide epidemic risk during the initial phase of the outbreak and prompt employment of a possibly large number of quarantines, if necessary. This strategy is particularly effective when a possibly large number of cheap, low-sensitivity rapid tests is available (72), as the prior information about the sensitivity of the tests can be included in the Bayesian probabilistic approach (39). Notice that a timely intervention ensures better containment in the long run, but also a lower time-integrated social cost (e.g. lower total number of isolated individuals). In this regard, a more in-depth study of probabilistic contact tracing strategies with intervention thresholds appears compelling.

In the numerical experiments, all contact tracing methods are either model-free (DCT, ICT, TTQ) or assume much simpler epidemic models (BP, SMF) compared to those used to generate the underlying epidemic traces. It follows that the superior performance shown by probabilistic contact tracing methods is not attributable to a greater knowledge of the real transmission mechanisms of the specific epidemics. This feature makes us believe that the results discussed in this work will remain consistent, at least qualitatively, in a real-world scenario with epidemic data from the necessarily much more complex diffusion in a human population, in the event of new variants or other emerging diseases with similar properties. Finally, these probabilistic contact tracing methods are sufficiently flexible to work with other prevention and mitigation measures, especially in the presence of vaccinated individuals and selective mobility restrictions.

Methods

Contact tracing methods

We report a brief description of the overall set of ranking techniques referring to the Supplementary Material for the implementation details.

Downloaded from https://academic.oup.com/pnasnexus/article/3/9/pgae377/7748389 by guest on 07 October 2024

- DCT. When individuals are tested positive, their recent contacts (within a one-week time window) are considered eligible for testing. When the number of individuals to be reached exceeds the number of available tests, we uniformly sample for testing as many of them as the maximum number of tests. This protocol is similar to the one published in Barrat et al. (35).
- ICT. Similarly to the probabilistic contact tracing technique, this method returns a score quantifying how likely each individual may be infected at the observation time. Exploiting both the positive and negative results of the tests, this method counts the number of potentially exposed events that each individual has had within a temporal window of one week. This type of potentially infectious contact occurs if (a) the considered individual has never been tested or has always been negatively tested in the past, and (b) the time of the contact lies in the time interval ranging from the last time the potential infector has been negatively tested (before being positively tested), and the first time it has been negatively tested after the infection (in case there is no such occurrence, this corresponds to the observation time). Here we have assumed that the process is irreversible, or, in other words, the time window we consider is sufficiently small to assume that, after a first infection, the acquired immunity preserves individuals from further infections after recovery.
- Simple Mean Field (SMF). This method assumes Markovian Susceptible-Infected-Recovered (SIR) dynamics with, when available to the app, heterogeneous infection probabilities mirroring, for instance, a diverse duration of the contacts. When an individual tests positive, SMF assumes that the infection occurred t_{SMF} days before (here and in Baker et al. (39) $t_{SMF} = 5$ as it seems to better fit the COVID-19 features). Finally, the SMF-based ranker estimates an approximated marginal probability of the state of all individuals at each time step of the dynamics. The values obtained at the observation time for the infected state are considered as a proxy for the individual risks. More details can be found in Ref. (39) and in the Supplementary Material.
- Belief Propagation (BP). Similarly to SMF, BP assumes that the underlying infection can be modeled as an SIR dynamic. Though, at difference with SMF, some intrinsic COVID-19 features are encoded in time-dependent infection and recovery rates resulting in a non-Markovian SIR model. Observations of the states of the individuals, i.e. the results of the medical tests, are properly introduced in the model by means of a Bayesian framework. This allows us to deal with imprecise test outcomes, mirroring the false negative and positive rates of the tests. Through the application of BP, the overall epidemic dynamics are reconstructed by inferring the infection and recovery times of all individuals. From this information, one can compute the individual probability of being infected at the observation time and, therefore, an estimate of the risk. The label $\tau = 7$ refers to the implementation used in Ref. (39), in which the risk is computed from the aggregate probability of infection and recovery times in a time window of $\tau = 7$ days. When a threshold is set, all individuals with a probability of being in the infected state larger than the threshold are tested. See Ref. (39) for a detailed description and Supplementary Material for the implementation details used for the StEM, Covasim, and OpenABM.
- Test-Trace-Quarantine (TTQ). This containment strategy is integrated into Covasim (69) and relies on the MCT process included in the model. This method traces individuals who have come in contact with confirmed infected ones (with a probability p_{trace} for each contact to be traced) and puts them in the so-called preemptive quarantine (PQ). In this state, which is unique to the Covasim model, individuals reduce their infectiousness levels. In the TTQ strategy, individuals are tested each day with a probability that depends both on their state (symptomatic or asymptomatic) and the time elapsed since their entrance into PQ (see Supplementary Material for details). As implemented in Covasim, this strategy does not limit the number of tests performed each day. To perform a fair comparison with the other containment techniques, in the regime of a limited number of tests, we adopted a modified version of the process, called TTQ-N, where the individuals to be tested are randomly chosen, drawing first from the set of symptomatic individuals and then with a probability proportional to the one used in TTQ. The process stops when the maximum number of tests is reached. Moreover, since individuals in the PQ state are subject to a reduction in transmission probability, in the results on Covasim shown in Figure 1, MCT is applied together with the other tracing methods in order to produce fair comparisons between TTQ and the other methods.

Agent-based models.

This section contains some important implementation details of the agent-based simulations. A brief description of the three models considered is reported in the Supplementary Material.

- OpenABM. The model introduced in Ref. (44) exploits discrete-time non-Markovian stochastic processes to simulate an epidemic spreading on an age-stratified population interacting on a multilayer synthetic graph, with demographic data based on the UK census (additional details are given in the Supplementary Material). The efficacy of probabilistic inference using BP and SMF against standard contact-tracing technique in the epidemic containment was already discussed in (39). Here we focus only on quantifying their performance w.r.t. the detection of super spreaders and forward/backward infections. The results presented in Sections and are obtained by simulating a population of $N = 10^5$ individuals for $T = 20$ days, with an initial number $N_{pz} = 100$ of infected individuals. All the other model parameters are not changed with respect to the default implementation of the simulator discussed in the original work. As the OpenABM model distinguishes between asymptomatic states and different classes of symptomatic ones (mild, severe), observations are performed on a daily basis on the full population of severe symptomatic and on 30% of mild symptomatic individuals, i.e. the same setting used in (39) for the online containment.
- Covasim. The work in Ref. (45) introduces Covasim, an agent-based model that includes country-specific demographic information such as age structure and population size. The contact networks used in Covasim comprise both an individual scale (these contacts are static) and a community scale (these interactions are randomly redrawn over time) to cope with households and social interactions. For this work, we use a population of 70, 000 individuals, with contact features matching those of the Seattle Metropolitan Area (as done in Ref. (45)). The epidemic model underlying the Covasim dynamic is a discrete-time non-Markovian process involving susceptible, exposed, several infectious states (an asymptomatic and presymptomatic state and three symptomatic states

Downloaded from https://academic.oup.com/pnasnexus/article/3/9/pgae377/7748389 by guest on 07 October 2024

to account for mild, severe, and critical conditions), as well as a recovered state. All individual transition times between these states are log-normally distributed. Special attention is devoted to the transmission of the disease; when a susceptible and an infectious individual meet, the transmission probability associated with this event depends on both individual viral-load-based transmissibility and susceptibility, and the social layer the contact belongs to. These ingredients favor the occurrence of superspreading events.

- StEM. The model proposed in Ref. (46) combines publicly available demographic data and automatic geo-referencing to produce continuous-time individual mobility traces with realistic features. In particular, we run mobility simulations on the urban area of Tübingen (Germany), having 90,546 individuals distributed in 47,309 houses. All the accessible venues fall into five categories (education, social places, public transport, offices, and supermarkets). Each inhabitant can visit a subset (one education venue, ten social places, five public transportation, one office, and two supermarkets) of the 1,487 available locations assigned with a probability that depends on the house-location distance. The duration of the visits depends on the location (2 hours at education, 1.5 hours at social places, 0.2 hours for public transport, 2 hours for working places, and 0.5 hours supermarket). Simulated mobility data is used to compute infectious contacts within a continuous-time non-Markovian stochastic model that includes an exposed state and multiple infected states to cope with asymptomatic, presymptomatic, and symptomatic individuals. Exposures depend on the state of the infectors, an exposure rate (set to 0.05 for all locations), and a kernel term that allows one to accommodate environmental transmissions. All contacts are available to the containment methods except those due to a small but continuous influx of untraceable exogenous exposures (as in the default setting, we set five of such events per 100,000 inhabitants and per week).

Acknowledgments

We are grateful to Lars Lorch for his help in managing the implementation of StEM. Computational resources were provided by HPC@POLITO, a project of Academic Computing within the Department of Control and Computer Engineering at the Politecnico di Torino (www.hpc.polito.it), and by the SmartData@PoliTO (smartdata.polito.it) interdepartmental center on Big Data and Data Science.

Supplementary Material

Supplementary material is available at *PNAS Nexus* online.

Funding

G.C. acknowledges support from the Comunidad de Madrid and the Complutense University of Madrid (UCM) through the Atracción de Talento programs (Refs. 2019-T1/TIC-13298). This study was carried out within the FAIR - Future Artificial Intelligence Research project and received funding from the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR)—Missione 4 Componente 2, Investimento 1.3—D.D. 1555 11/10/2022, PE00000013). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Author Contributions

A.P.M. designed the research, performed simulations about the epidemic containment, performed static analysis on the transmission features, analyzed the results and wrote the manuscript. F.M. performed simulations about the epidemic containment, performed static analysis on the transmission features, analyzed the results and wrote the manuscript. A.B. designed the research, analyzed the results and wrote the manuscript. G.C. performed static analysis on the transmission features, analyzed the results, and wrote the manuscript. L.D.A. designed the research, analyzed the results, and wrote the manuscript.

Preprints

A preprint of this article is published at [arXiv.2312.00910].

Data Availability

All data, codes, and materials used for producing these results are available at (73). All other data are present in the paper and the Supplementary Material.

References

1 Bonaccorsi G, *et al.* 2020. Economic and social consequences of human mobility restrictions under COVID-19. *Proc Natl Acad Sci U S A*. 117(27):15530–15535.
2 Kretzschmar ME, *et al.* 2022. Challenges for modelling interventions for future pandemics. *Epidemics*. 38:100546.
3 Perra N. 2021. Non-pharmaceutical interventions during the COVID-19 pandemic: a review. *Phys Rep*. 913:1–52.
4 Hellewell J, *et al.* 2020. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob Health*. 8(4):e488–e496.
5 Bi Q, *et al.* 2020. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect Dis*. 20(8):911–919.
6 Lavezzo E, *et al.* 2020. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of vo'. *Nature*. 584(7821):425–429.
7 Firth JA, *et al.* 2020. Using a real-world network to model localized COVID-19 control strategies. *Nat Med*. 26(10):1616–1622.
8 Keeling MJ, Hollingsworth TD, Read JM. 2020. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). *J Epidemiol Community Health*. 74(10):861–866.
9 Mastrandrea R, Fournet J, Barrat A. 2015. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS One*. 10(9):e0136497.
10 Smieszek T, *et al.* 2016. Contact diaries versus wearable proximity sensors in measuring contact patterns at a conference: method comparison and participants' attitudes. *BMC Infect Dis*. 16(1):1–14.
11 Danquah LO, *et al.* 2019. Use of a mobile application for Ebola contact tracing and monitoring in northern Sierra Leone: a proof-of-concept study. *BMC Infect Dis*. 19(1):1–12.
12 Sacks JA, *et al.* 2015. Introduction of mobile health tools to support Ebola surveillance and contact tracing in Guinea. *Glob Health Sci Pract*. 3(4):646–659.
13 Ferretti L, *et al.* 2020. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*. 368(6491):eabb6936.

Downloaded from https://academic.oup.com/pnasnexus/article/3/9/pgae377/7748389 by guest on 07 October 2024

14 Kucharski AJ, *et al.* 2020. Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *Lancet Infect Dis*. 20(10):1151–1160.
15 Braithwaite I, Callender T, Bullock M, Aldridge RW. 2020. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. *Lancet Digit Health*. 2(11):e607–e621.
16 Chien L-C, Beÿ CK, Koenig KL. 2022. Taiwan's successful COVID-19 mitigation and containment strategy: achieving quasi population immunity. *Disaster Med Public Health Prep*. 16(2): 434–437.
17 Oh J, *et al.* 2020. National response to COVID-19 in the Republic of Korea and lessons learned for other countries. *Health Syst Reform*. 6(1):e1753464.
18 Aslam H, Hussain R. 2020. *Fighting covid-19: lessons from china, south korea and japan*. Sustainable Development Policy Institute.
19 Huang Z, *et al.* 2020. Performance of digital contact tracing tools for COVID-19 response in Singapore: cross-sectional study. *JMIR Mhealth Uhealth*. 8(10):e23148.
20 Amann J, Sleigh J, Vayena E. 2021. Digital contact-tracing during the COVID-19 pandemic: an analysis of newspaper coverage in Germany, Austria, and Switzerland. *PLoS One*. 16(2):e0246524.
21 Bengio Y, *et al.* 2020. The need for privacy with public digital contact tracing during the COVID-19 pandemic. *Lancet Digit Health*. 2(7):e342–e344.
22 Mello MM, Wang CJ. 2020. Ethics and governance for digital disease surveillance. *Science*. 368(6494):951–954.
23 Jacob S, Lawarée J. 2021. The adoption of contact tracing applications of COVID-19 by European governments. *Policy Des Pract*. 4(1):44–58.
24 Munzert S, Selb P, Gohdes A, Stoetzer LF, Lowe W. 2021. Tracking and promoting the usage of a COVID-19 contact tracing app. *Nat Hum Behav*. 5(2):247–255.
25 Aarogya setu app. 2020. https://www.aarogyasetu.gov.in/.
26 Nhs COVID-19 app. 2020. https://covid19.nhs.uk/.
27 Bay J, *et al.* 2020. Bluetrace: a privacy-preserving protocol for community-driven contact tracing across borders. Government Technology Agency-Singapore, Tech. Rep. 18.
28 Apple and Google. 2020. Privacy-preserving contact tracing. https://covid19.apple.com/contacttracing.
29 Chan J, *et al.* 2020. PACT: privacy-sensitive protocols and mechanisms for mobile contact tracing. *IEEE Data Eng. Bull*. 43(2):15–35.
30 Troncoso C, *et al.* 2022. Deploying decentralized, privacy-preserving proximity tracing. *Commun. ACM*. https://doi.org/10.1145/3524107.
31 Kendall M, *et al.* 2020. Epidemiological changes on the isle of wight after the launch of the NHS test and trace programme: a preliminary analysis. *Lancet Digit Health*. 2(12):e658–e666.
32 Rodríguez P, *et al.* 2021. A population-based controlled experiment assessing the epidemiological impact of digital contact tracing. *Nat Commun*. 12(1):1–6.
33 Salathé M, *et al.* 2020. Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland. *Swiss Med Wkly*. 150(5153):w20457. https://doi.org/10.4414/smw.2020.20457.
34 Wymant C, *et al.* 2021. The epidemiological impact of the NHS COVID-19 app. *Nature*. 594(7863):408–412.
35 Barrat A, Cattuto C, Kivelä M, Lehmann S, Saramäki J. 2021. Effect of manual and digital contact tracing on COVID-19 outbreaks: a study on empirical contact data. *J R Soc Interface*. 18(178): 20201000.
36 Cencetti G, *et al.* 2021. Digital proximity tracing on empirical contact networks for pandemic control. *Nat Commun*. 12(1):1–12.
37 Contreras S, *et al.* 2021. The challenges of containing SARS-CoV-2 via test-trace-and-isolate. *Nat Commun*. 12(1):1–13.
38 Alsdurf H, *et al.* 2020. Covi white paper, arXiv, arXiv:2005.08502, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.2005.08502.
39 Baker A, *et al.* 2021. Epidemic mitigation by statistical inference from contact tracing data. *Proc Natl Acad Sci U S A*. 118(32).
40 Fenton N, *et al.* 2020. A Bayesian network model for personalised COVID19 risk assessment and contact tracing. https://doi.org/10.1101/2020.07.15.20154286.
41 Murphy K, Kumar A, Serghiou S. 2021. Risk score learning for COVID-19 contact tracing apps. In: Machine Learning for Healthcare Conference. PMLR. p. 373–390.
42 Altarelli F, Braunstein A, Dall'Asta L, Lage-Castellanos A, Zecchina R. 2014. Bayesian inference of epidemics on networks via belief propagation. *Phys Rev Lett*. 112(11):118701.
43 Braunstein A, Ingrosso A. 2016. Inference of causality in epidemics on temporal contact networks. *Sci Rep*. 6(1):27538.
44 Hinch R, *et al.* 2021. OpenABM-Covid19-an agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. *PLoS Comput Biol*. 17(7):e1009146.
45 Kerr CC, *et al.* 2021. Covasim: an agent-based model of COVID-19 dynamics and interventions. *PLoS Comput Biol*. 17(7):e1009149.
46 Lorch L, *et al.* 2022. Quantifying the effects of contact tracing, testing, and containment measures in the presence of infection hotspots. *ACM Trans Spatial Algo Syst*. 8(4):1–28.
47 Kojaku S, Hébert-Dufresne L, Mones E, Lehmann S, Ahn Y-Y. 2021. The effectiveness of backward contact tracing in networks. *Nat Phys*. 17(5):652–658.
48 Bradshaw WJ, Alley EC, Huggins JH, Lloyd AL, Esvelt KM. 2021. Bidirectional contact tracing could dramatically improve COVID-19 control. *Nat Commun*. 12(1):1–9.
49 Tufekci Z. 2020. This overlooked variable is the key to The Pandemic. section: health.
50 Althouse BM, *et al.* 2020. Superspreading events in the transmission dynamics of SARS-CoV-2: opportunities for interventions and control. *PLoS Biol*. 18(11):e3000897.
51 Endo A, *et al.* 2020. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res*. 5.
52 Lau MSY, *et al.* 2020. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. *Proc Natl Acad Sci U S A*. 117(36):22430–22435.
53 Lemieux JE, *et al.* 2021. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science*. 371(6529):eabe3261.
54 Stein RA. 2011. Super-spreaders in infectious diseases. *Int J Infect Dis*. 15(8):e510–e513.
55 Sun K, *et al.* 2021. Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science*. 371(6526):eabe2424.
56 Paredes MI, *et al.* 2024. Underdetected dispersal and extensive local transmission drove the 2022 mpox epidemic. *Cell*. 187(6): 1374–1386.
57 Smith J. 2022. Yes, it matters who is spreading monkeypox. *Lancet Infect Dis*. 22(10):1416–1417.
58 Ward T, *et al.* 2024. Understanding the infection severity and epidemiological characteristics of mpox in the UK. *Nat Commun*. 15(1):2199.
59 Oshitani H, *et al.* 2020. Cluster-based approach to coronavirus disease 2019 (COVID-19) response in Japan–February–April 2020. *Jpn J Infect Dis*. 73(6):491–493.

Downloaded from https://academic.oup.com/pnasnexus/article/3/9/pgae377/7748389 by guest on 07 October 2024

60 Lee SW, *et al.* 2020. Nationwide results of COVID-19 contact tracing in South Korea: individual participant data from an epidemiological survey. *JMIR Med Inform*. 8(8):e20992.
61 Taylor L. 2020. Uruguay is winning against COVID-19. this is how. *BMJ*. 370.
62 Endo A, *et al.* 2020. Implication of backward contact tracing in the presence of overdispersed transmission in COVID-19 outbreaks. *Wellcome Open Res*. 5.
63 Chang S, *et al.* 2021. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*. 589(7840):82–87.
64 Flaxman S, *et al.* 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*. 584(7820):257–261.
65 Pullano G, *et al.* 2021. Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature*. 590(7844):134–139.
66 Aleta A, *et al.* 2020. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat Hum Behav*. 4(9):964–971.
67 Gupta P, *et al.* 2020. Covi-agentsim: an agent-based model for evaluating methods of digital contact tracing, arXiv, arXiv:2010.16004, preprint: not peer reviewed. https://doi.org/10.48550/arXiv.2010.16004.
68 Lasser J, *et al.* 2022. Assessing the impact of SARS-CoV-2 prevention measures in Austrian schools using agent-based simulations and cluster tracing data. *Nat Commun*. 13(1):1–17.
69 Kerr CC, *et al.* 2021. Controlling COVID-19 via test-trace-quarantine. *Nat Commun*. 12(1):2993.
70 Dinnes J, *et al.* 2022. Rapid, point-of-care antigen tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Syst Rev*. (7). https://doi.org//10.1002/14651858.CD013705.pub3.
71 Wong F, Collins JJ. 2020. Evidence that coronavirus superspreading is fat-tailed. *Proc Natl Acad Sci U S A*. 117(47):29416–29418.
72 Kennedy-Shaffer L, Baym M, Hanage WP. 2021. Perfect as the enemy of good: tracing transmissions with low-sensitivity tests to mitigate SARS-CoV-2 outbreaks. *Lancet Microbe*. 2(5):e219–e224.
73 Sibyl Team. Sib. 2020. https://github.com/sibyl-team/.

Downloaded from https://academic.oup.com/pnasnexus/article/3/9/pgae377/7748389 by guest on 07 October 2024