Epidemic spreading and digital contact tracing: Effects of heterogeneous mixing and quarantine failures

Abbas K. Rizi [1], Ali Faqeeh,[1,2,3] Arash Badie-Modiri [1], and Mikko Kivelä [1]

[1]*Department of Computer Science, School of Science, Aalto University, FI-00076, Finland*
[2]*Mathematics Applications Consortium for Science & Industry, University of Limerick, Limerick V94 T9PX, Ireland*
[3]*Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, Indiana 47408, USA*

Contact tracing via digital tracking applications installed on mobile phones is an important tool for controlling epidemic spreading. Its effectivity can be quantified by modifying the standard methodology for analyzing percolation and connectivity of contact networks. We apply this framework to networks with varying degree distributions, numbers of application users, and probabilities of quarantine failures. Further, we study structured populations with homophily and heterophily and the possibility of degree-targeted application distribution. Our results are based on a combination of explicit simulations and mean-field analysis. They indicate that there can be major differences in the epidemic size and epidemic probabilities which are equivalent in the normal susceptible-infectious-recovered (SIR) processes. Further, degree heterogeneity is seen to be especially important for the epidemic threshold but not as much for the epidemic size. The probability that tracing leads to quarantines is not as important as the application adoption rate. Finally, both strong homophily and especially heterophily with regard to application adoption can be detrimental. Overall, epidemic dynamics are very sensitive to all of the parameter values we tested out, which makes the problem of estimating the effect of digital contact tracing an inherently multidimensional problem.

I. INTRODUCTION

Until effective vaccines are widely deployed in a pandemic era, carefully timed nonpharmaceutical interventions [1] such as wearing face masks [2], school closures, travel restrictions, and contact tracing [3–7] are the best tools we have for curbing the pandemic. Contact tracing is an attempt to discover and isolate asymptomatic or presymptomatic (exposed) individuals. In the absence of herd immunity, contact tracing is a potent low-cost intervention method since it puts people into quarantine where and when the disease spreads. Therefore, it can have a significant role in containing a pandemic by relaxing social-distancing interventions [8], providing an acceptable trade-off between public health and economic objectives [9,10], developing sustainable exit strategies [11,12], identifying future outbreaks [13], and reaching the "source" of infection [14].

Thanks to the emergence of low-cost wearable health devices [15–22] and mobile software applications, digital contact tracing can now be deployed with higher precision without the problems of manual contact tracing, such as the tracing being slow and labor intensive or people's hesitation to give identifying data about their contacts due to blame, fear, confusion, or politics. On the other hand, smartphones and wearable devices also offer continuous. access to real-time physiological data, which can be used to tune other nonpharmaceutical or pharmaceutical strategies. Modern apps enable us to monitor COVID-19 symptoms [23–25]; identify its hotspots [26]; track mosquito-borne diseases such as Malaria, Zika, and Dengue [27,28]; and detect microscopic pathogens.

In both forms—manual [4,5,29–37] and digital [38–44]—contact tracing has been commonly considered as an effective strategy and different empirical data sets have validated this claim in short-time population-based controlled experiments [38,40]. It has been estimated that for every percentage point increase in app-users, the number of cases can be reduced by 2.3% (in statistical analysis) [45]. However, such a linear view of the benefits of the app usage is likely too simplistic and ignores the complexities disease spreading, especially in heterogeneous populations [46–49]. For instance, degree-heterogeneity in the contact network [50] can alter epidemiological properties in the form of variance in final outbreak size [51], vanishing epidemic threshold [49,52], hierarchical spreading [53], strong finite-size effects [54], and universality classes for critical exponents [55]. Moreover, the existence of superspreaders dictates the extent to which a virus spreads in a bursty fashion [56–58], especially when there is high individual-level variation in the number of secondary transmissions [53,59,60]. Therefore, to evaluate the effectiveness of contact-tracing, degree-heterogeneity and app adoption of superspreaders [61,62] should be taken into account. Note that in some parameter settings, contact tracing may not be effective enough [8,63,64].

A potentially important factor in the effectiveness of the contact tracing apps is related to how the app-using and non-app-using populations are mixed. Several studies have shown that similar people with similar features are more likely to be in contact with each other than with people with different types of features. This phenomenon is known as homophily [65–67]. It has been reported in app adoption directly [40]

and indirectly through correlation in app adoption and other features exhibiting homophily, such as jobs, age, income, and nationality [68–70]. Therefore, the fraction of population that adopts the app is not the only important factor for reducing the peak and total size of the epidemic but also the amount of homophily in app adoption can potentially have a significant role.

Since the World Health Organization has declared the COVID-19 outbreak as a Public Health Emergency of International Concern, network scientists have developed different approaches toward analyzing epidemic tracing and mitigation with apps. Using the toolbox of network science, different groups have investigated the effectiveness of contact tracing based on the topology and directionality of contact networks [14,44,71–77]. Recently, a mathematical framework aimed at understanding how homophily in health behavior shapes the dynamics of epidemics has been introduced by Burgio *et al.* [78]. This study expanded the model of Bianconi *et al.* [71] and computed the reproduction number and attack rate in a homophilic population using mean-field equations.

Our study investigates the effect of varying app coverage on the epidemic's threshold, probability, and expected size in homogeneous and heterogeneous contact networks with and without homophily or heterophily in app adoption. Further, we explore the effect of distributing the apps randomly and preferentially to high-degree nodes [71] in these scenarios. Our main focus is on the epidemic threshold and the final size of the epidemics. Therefore, we assume the dynamics of the epidemic to be governed by the simple SIR model [50]. This model can be easily mapped to a static bond-percolation problem [79,80] so that the epidemiological properties can be measured based on the topological structure of the underlying network [50,73,81–84]. Note that, more complex disease transmission models, such as SEIR models in which there is an infected-but-not-contagious period E, are also covered by this formalism [79,85]. The difference in the spreading framework with the app to the normal one is that the infection cannot spread further if it passes a link between two app-users (app-adopters). That is, the infection process model needs to include the memory of the type of node it is coming from. We then extend the percolation framework such that we can add memory [86,87] to it in order to keep track of the infection path. This leads to the observation that the epidemic size is not the same as the epidemic probability as it would be in this model without the app-users [88].

Our results are largely based on mean-field-type calculations of the percolation problem, which are confirmed by explicit simulations of SIR epidemic process and measurements of component sizes in finite networks. Our findings show that (1) the number of app-users has a direct effect on the epidemic size and epidemic probability and the difference between these two observables is larger in high-degree targeting strategy; (2) epidemics can be controlled to a much better degree in the high-degree targeting strategy; (3) even though degree-heterogeneity can strongly affect or even eliminate the epidemic threshold, high-degree targeting strategy can compensate this effect and increase the threshold significantly; (4) increasing heterophily from random mixing always increases the outbreak size and lowers the epidemic threshold; (5) increasing homophily does the opposite until an optimum,

that is below the maximum homophily case, is reached; and, finally, (6) the probability of contact tracing succeeding in preventing further infections is not as crucial as the fraction of app-users, but can still have significant effects on the epidemic size and epidemic threshold. The only exception is when the apps are distributed to heterogeneous networks with the high-degree targeting strategy.

II. MODELLING APPROACH

A. Disease model and connection to percolation

We employ a SIR disease model on networks with additional dynamics given by the disease interactions in the presence of the disease tracking application. In the model, without the tracking application, an infected (I) node will eventually infect a neighboring susceptible (S) node with a transmission probability p independently of other infections. The simulations are performed with a model where the infected nodes try to infect their susceptible neighbors with independent Poisson processes with rate β and go to the removed state (R) after fixed time τ. The fixed recovery time ensures that every infected individual, regardless of app adoption, can infect a susceptible neighbor independently with a transmission probability $p = 1 - e^{-\beta\tau}$ [79,89]. These assumptions allow us to study the SIR processes using component size distributions of undirected networks where parts of the links are randomly removed [79,85,88–90]: An epidemic starting from a single node can reach any other node exactly when there is a path of such transmitting links connecting them, i.e., they are in the same component in a network where the potential contact links are removed with probability p. Thus, the epidemic threshold, epidemic probability and epidemic size can be read from percolation simulations [79,85,88–90] (see Sec. II B). Note that without fixed recovery time, the presence of spreading paths through neighboring links would not be independent, and this would not be a bond percolation problem in an undirected network where edges are removed independently. However, the epidemic threshold, final epidemic size, and the expected outbreak size below the epidemic threshold would still be correctly predicted by this methodology [88,89].

We model the effect of applications to the disease spreading as follows: If an app-user infects another app-user, then that second node will get infected but will quarantine themselves with probability p_{app}. The quarantined user will have no further connections that would spread the infection they received from the other app-user. A substantial deviation from a realistic spreading case in our model is that the quarantine does not prevent the disease spreading to the quarantined node through a third node. That is, we only model the primary infection path from the other app-user causing the alarm but do not stop the possible concurrent secondary infection paths from a third node. Strictly speaking, this simplification in the modeling returns a lower bound on the effectiveness of the app-based contact tracing, but given that our contact network models are sparse random graphs (see Sec. II C) that do not contain local loops, the difference can only be observed if a large enough fraction of the population is infected at the same time. Critically, this does not affect the epidemic threshold
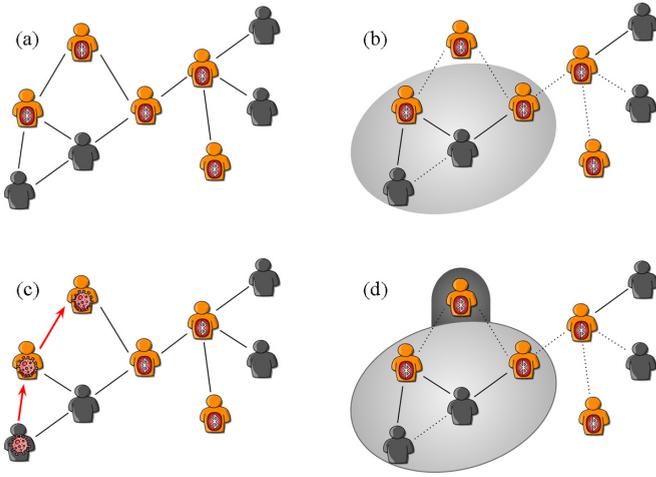
FIG. 1. (a) Original contact network with app-users marked with the oval symbol. (b) The normal largest component, after the dotted links have been removed in the percolation process at random. When apps are working perfectly, links between a pair of app-users are removed with probability $p_{app} = 1$ and other links are removed with probability p. (c) An example for a path of infection: The second app-user can be infected; therefore, it must be included in the outbreak size. (d) Extending the giant component to include the secondary app infections. The second infected app-adopter is added to the giant component with transmission probability p.

but could have implications for parameter regions where the epidemic size is large, depending on the quarantine durations.

The SIR spreading process can be mapped to a slightly more complicated percolation problem in the presence of apps [44,71]. To model app-user quarantines, one needs to delete the links between two app-users with the probability of successful quarantine due to the app, p_{app}. This ensures that we ignore the infection paths through two app-users when one of them is successfully quarantined. However, removing these links also removes the second app-user from the component, even though they are infected. To correct this, we need first to find the network components and then extend them by including all app-users outside of the component connected to another app-user (and considering the probability p that the link is kept). See Fig. 1 for an illustration of this process, which leads us to two definitions of components: normal and extended.

B. Components, epidemic size, and epidemic probability

In the SIR model without apps, the component size distribution can be used to describe the late stages of the epidemics approximately. Given an initially infected node, the size of the component it belongs to determines the size of the outbreak. In an infinitely large population, we say that an outbreak is an epidemic if it spans a nonzero fraction of the population. The relationship between percolation and the final epidemic size is straightforward if the population is large enough that it can be approximated with an infinite undirected transmission network [79,88]. In this case, the percolation threshold gives the epidemic threshold and below it, an outbreak always spans only a zero fraction of the population because all the components are of finite size. Above the percolation threshold,

there is a single giant component that spans $s_{max} = S_{max}/N$ fraction of the nodes. This is equivalent to both the size of the epidemic, given that there is one, and the probability that there is an epidemic starting from a single initially infected node [79,88]; s_{max} is the fraction of nodes that can be reached from the giant component (epidemic size) and the probability that randomly chosen node belongs to the giant component (probability of the epidemic). The expected epidemics size in a fraction of eventually infected nodes is, in this case, given as s_{max}^2.

When we introduce apps to the spreading process, the equivalence of the epidemic size and epidemic probability breaks down. Both the normal component and the extended component become important. The component size still gives the probability that there is an epidemic, as is the case without the apps. However, the epidemic size, given that there is one, is now given by the extended component size s'_{max}. The expected epidemic size is then given by $s_{max} s'_{max}$.

Similar relationships hold for finite-size systems. For example, the expected size of the epidemics from single source becomes

$$\langle E \rangle = \sum_c \frac{S_c}{N} S'_c, \qquad (1)$$

where S_c is the normal size and S'_c is the extended size of the component c and N is the total number of nodes. In this formula, S_c/N gives the probability that the initially infected node is in the component c and S'_c gives the size of the epidemic if a node in component c is chosen.

C. Network models

We aim to study how the network topology, amount, and distribution of app-users over the network affect the epidemics. We study networks with degree distribution $P(k)$ and average degree $\langle k \rangle$ such that each node is an app-user with probability π_a and not an app-user with probability $1 - \pi_a$. We distribute the app-users with one of two strategies: (1) uniformly at random or (2) by distributing the apps in the order of their degree such that the high-degree nodes get the apps first.

We use three different models to generate the network topology. We use (i) Poisson or Erdős-Rényi (ER) random graphs [91] to model homogeneous contact patterns and (ii) scale-free networks generated with the Chung-Lu model (CL) [92,93] to model heterogeneous networks. In generation of CL networks, the expected degree of each node is drawn from a continuous power-law distribution $P(k) \propto k^{-3}$ such that the minimum expected degree is set to a value that gives us the expected average degree $\langle k \rangle$ of our choice. Given a sequence of expected degrees $W = \{w_1, w_2, \ldots, w_n\}$, Miller algorithm [94] assigns a link between node u and node v with probability $p_{uv} \propto w_u w_v$. This algorithm returns a network without multiple links with almost the same power-law degree distribution.

We model homophily (and heterophily) with regards to apps usage with (iii) a modular network model (MN) introduced in Refs. [95,96] with two groups of nodes: app-users and non-app-users. This model starts with a degree sequence produced either by the ER or CL models and connects the nodes depending on which groups they belong with

probabilities π_{aa} (app-user to app-user), π_{an} (app-user to non-app-user), π_{na} (non-app-user to app user), and π_{nn} (non-app-user to non-app-user). We only need to fix one of these probabilities, π_{aa}, and other types of links are formed with probabilities $\pi_{an} = 1 - \pi_{aa}$, $\pi_{na} = \frac{\pi_a}{1-\pi_a}(1 - \pi_{aa})$, and $\pi_{nn} = 1 - \pi_{na} = \frac{1-\pi_a-\pi_a(1-\pi_{aa})}{1-\pi_a}$, where π_a is the probability that a person is an app-user and the second equality comes from the balance between the number of links from app-users to non-app-users and from non-app to app-users, that is, $\pi_a N \pi_{an} \langle k \rangle = (1 - \pi_a) N \pi_{na} \langle k \rangle$. The numerical simulations of the MN work by randomly choosing a group for half edges with the given probabilities and matching them to each other uniformly randomly. This can lead to self-links and multilinks, which these are discarded after the randomization procedure.

While there is no correlation between the app adoption status in homogeneous (ER) or heterogeneous (CL) networks above, in the third model (MN), the existence of homophily or heterophily of the network structure is determined by comparing π_{aa} to its value for the neutral case with no homophily or heterophily. In the absence of homophily or heterophily, $\pi_{aa} = \eta_a$, where η_a is the ratio of the number of links that emerge from app-users to the total degree; this is because if the nodes were connected purely at random, then the probability that a link from an app-user connects it to another app-user equals the ratio of the number of stubs that app-users have to the total number of stubs, i.e., η_a. In the case of a random selection of app-users $\eta_a = \pi_a$, since both app-users and non-app-users have on average the same number of stubs and the fraction of stubs that app-users have equals the fraction of app-users in the system, i.e., π_a. In a high-degree targeting strategy, the number of stubs that app-users have on average is larger than that of non-app-users. In that case, η_a can be calculated from the degree distribution (see Sec. III A). When $\pi_{aa} > \eta_a$, app-users are more likely to be connected to each other than in a network in which a fraction of η_a of them being uniformly randomly placed. On the other hand, when $\pi_{aa} < \eta_a$ nodes are more likely to be connected to the nodes of the other type (heterophilic network). In the heterophilic regime, for some pairs of (π_{aa}, π_a), networks are not realizable because of the constraints explained in Sec. III A. The white region in Fig. 6 shows the region of π_{aa}-π_a plane that networks cannot be created in that parameter space.

III. ANALYTIC AND SIMULATION METHODS

The epidemics are studied here with various methods of approximation. We employ analytical computations based on mean-field-type approximations to efficiently analyze our models' wide parameter space and provide explicit formulas for our main observable quantities. Here an approximation based on branching processes [97] can be used to determine the critical point. Following Ref. [44], a more detailed calculation based on percolation arguments will give us the component sizes which can be related to the final epidemic size and epidemic probability. Simulations of the network connectivity then complement these mean-field approximations. Finally, we explore the accuracy of the mean-field approximations via explicit simulations of the SIR model.

A. Giant component size from consistency equations

To study the behavior of the epidemic dynamics, we form consistency equations for the giant component size. In Ref. [44] the governing equations for the size of the epidemic and the transition point were obtained for the case of random networks in the absence of homophily. Here we derive the analytical results for the more general case of the spectrum of heterophilic to homophilic networks, a special case of which is the nonhomophilic networks of Ref. [44]. We consider that app-users and non-app-users might be connected together with a pattern different from pure random chance using the MN model.

We aim to write the self-consistent equations for the probability, u_n, that following a link to a non-app-user does not lead to the giant component and probability u_a, that following a link to an app-user does not lead to the giant component. Using these probabilities, the relative size of the giant component s and the relative size of the extended giant component s' can be obtained, where s is, in fact, the fraction of nodes infected through non-app-users, while s' also includes individuals who caught the infection through an app-user before they could quarantine themselves (see Sec. III C 1).

We need to know the probability u_n (u_a), that a randomly chosen link leading to a non-app-user (app-user) is not in the giant component. The probability that a non-app-user (app-user) is not connected to the giant component via a particular neighboring node is equal to the probability that *that* non-app-user (app-user) is not connected to the giant component via any of its other neighbors. A non-app-user is connected to another non-app-user with probability $\pi_{nn} = 1 - \pi_{na}$ and to an app-user with probability π_{na}. So, a link leading out from a non-app-user does not lead to the giant component if it leads to another non-app-user that is not in the giant component [which happens with probability $(1 - \pi_{na})u_n$] or an app-user that is not in the giant component (which happens with probability $\pi_{na}u_a$). That is, the total probability for following a link out from a non-app-user not leading to the giant component is $u_{n\rightarrow} = (1 - \pi_{na})u_n + \pi_{na}u_a$. Since the degree of neighboring nodes is disturbed according to the excess degree distribution q_k, the probability that a non-app-user that is encountered by following a link to it is not connected to the giant component via any of its k neighbors is $\sum_k q_k u_{n\rightarrow}^k$. This probability is, by definition, u_n, leading to the self-consistent equation below for u_n:

$$u_n = g_1[(1 - \pi_{na})u_n + \pi_{na}u_a], \qquad (2)$$

where g_1 is the generating function for excess degree distribution [50]. To find u_a, we can use the same treatment, except that we should consider how app-app connections depend on the probability of success in contact tracing [44]. If p_{app} is the probability the apps work as expected, then $1 - p_{app}$ is the probability that the app-user does not effectively quarantine after being been in contact by an infectious app-user. Therefore, u_a can be expressed as the self-consistent equation below:

$$u_a = g_1\{(1 - \pi_{aa})u_n + \pi_{aa}[p_{app} + (1 - p_{app})u_a]\}. \qquad (3)$$

Note that π_{na} is determined by the free parameters π_a and π_{aa} as we already showed that $\pi_{na} = \frac{\pi_a}{1-\pi_a}(1 - \pi_{aa})$.

Given u_n and u_a, the average probability that a node belongs to the giant component, or equivalently the fraction of the network occupied by the giant component, is now given by:

$$s = 1 - (1 - \pi_a)g_0[(1 - \pi_{\mathrm{na}})u_n + \pi_{\mathrm{na}}u_a]$$
$$- \pi_a g_0\{(1 - \pi_{\mathrm{aa}})u_n + \pi_{\mathrm{aa}}[p_{\mathrm{app}} + (1 - p_{\mathrm{app}})u_a]\}, \quad (4)$$

where g_0 is the generating function for degree distribution. We can approximate s' by writing:

$$s' = 1 - (1 - \pi_a)g_0[(1 - \pi_{\mathrm{na}})u_n + \pi_{\mathrm{na}}u_a]$$
$$- \pi_a g_0[(1 - \pi_{\mathrm{aa}})u_n + \pi_{\mathrm{aa}}u_a)], \quad (5)$$

where, as opposed to Eq. (4), the third term is not a function of p_{app} and the reason is that Eq. (4) assumes that if the app works (which happens with probability p_{app}) then the probability that a link connected to an app-user does not lead to the giant component is 1 (while if the app does not work it is u_a). However, whether the app works or not, the probability that an app-user does not get infected from another app-user is u_a. When apps work, if the second app-user is infected, she quarantines herself and does not infect any other node).

In the case of including a transmission probability p which is less than 1 (in the above equations it was assumed the links are transmitting with probability 1), Eqs. (2) and (3) will change to:

$$u_n = 1 - p + pg_1[(1 - \pi_{\mathrm{na}})u_n + \pi_{\mathrm{na}}u_a], \quad (6)$$

$$u_a = 1 - p + pg_1\{(1 - \pi_{\mathrm{aa}})u_n + \pi_{\mathrm{aa}}[p_{\mathrm{app}} + (1 - p_{\mathrm{app}})u_a]\}. \quad (7)$$

When the fraction π_a of nodes selected to adopt the app are all the highest degree nodes in the network, these nodes all have a degree higher than $k_a - 1$ such that they include some of k_a nodes and the rest are comprised of all nodes with degree larger than k_a. Then for the fraction η_a of the links protruding from the app-users (which are the top π_a fraction of nodes) we can write:

$$\eta_a = r^* k_a p_{k_a}/\langle k \rangle + \sum_{k_a+1}^{\infty} k p_k/\langle k \rangle, \quad (8)$$

$$= \sum_{k_{a,\mathrm{right}}}^{\infty} k p_k/\langle k \rangle, \quad (9)$$

where r^* is the fraction of degree k_a nodes that are app-users and in Eq. (9) we absorbed r^* into p_k so that $p_{k_{a,\mathrm{right}}} = r^* p_{k_a}$ represents the fraction of nodes in the network that have degree k_a and are app-users [so in Eq. (9), $k_{a,\mathrm{right}}$ takes the value k_a].

Then for a network with homo- or heterophily:

$$u_n = 1 - p + \frac{p}{1 - \eta_a} \sum_{k=0}^{k_{a,\mathrm{left}}} q_k[(1 - \pi_{\mathrm{na}})u_n + \pi_{\mathrm{na}}u_a]^k, \quad (10)$$

$$u_a = 1 - p + \frac{p}{\eta_a} \sum_{k_{a,\mathrm{right}}}^{\infty} q_k\{(1 - \pi_{\mathrm{aa}})u_n$$
$$+ \pi_{\mathrm{aa}}[p_{\mathrm{app}} + (1 - p_{\mathrm{app}})u_a]\}^k, \quad (11)$$

and

$$s = 1 - \sum_{k=0}^{k_{a,\mathrm{left}}} p_k[(1 - \pi_{\mathrm{na}})u_n + \pi_{\mathrm{na}}u_a]^k$$
$$- \sum_{k_{a,\mathrm{right}}}^{\infty} p_k\{(1 - \pi_{\mathrm{aa}})u_n + \pi_{\mathrm{aa}}[p_{\mathrm{app}} + (1 - p_{\mathrm{app}})u_a]\}^k.$$

$$(12)$$

A special case of which are networks with neutral (nonexisting) homophily, where π_{aa} is obtained to be equal to η_a and accordingly $\pi_{\mathrm{na}} = \eta_a$, therefore,

$$u_n = 1 - p + p\frac{1}{1 - \eta_a} \sum_{k=0}^{k_{a,\mathrm{left}}} q_k[(1 - \eta_a)u_n + \eta_a u_a]^k, \quad (13)$$

$$u_a = 1 - p + p\frac{1}{\eta_a} \sum_{k_{a,\mathrm{right}}}^{\infty} q_k\{\eta_a[p_{\mathrm{app}} + (1 - p_{\mathrm{app}})u_a]$$
$$+ (1 - \eta_a)u_n\}^k, \quad (14)$$

and

$$s = 1 - \sum_{k=0}^{k_{a,\mathrm{left}}} p_k[(1 - \eta_a)u_n + \eta_a u_a]^k$$
$$- \sum_{k_{a,\mathrm{right}}}^{\infty} p_k\{\eta_a[p_{\mathrm{app}} + (1 - p_{\mathrm{app}})u_a] + (1 - \eta_a)u_n\}^k. \quad (15)$$

These results predict the behavior of the epidemic dynamics in the thermodynamic limit. Therefore they describe the dynamics very well when the network size is large enough.

B. Mean-field approximation for the branching process

An alternative to writing the consistency equations for the giant component size is to assume that a branching process governs the epidemic dynamics. Then, a straightforward way of finding the epidemic threshold in the SIR model is to find the critical point of a branching process, where the branching factor is given by the expected excess degree q. In the epidemic setting, the branching factor $\bar{k}_e = pq$ gives the expected number of people one infected person infects during the epidemic process. Note that the branching factor has been used as the definition of the basic reproduction number R_0 [88], but is different from the basic reproduction number when it is defined in the networks as $R_0 = \beta/\gamma\langle k \rangle$ [80]. In the SIR model with the app, we need to duplicate the populations so that we separately track the ones without the app (S_n, I_n, and R_n) and with the app (S_a, I_a and R_a).

Given that the apps are uniformly distributed to π_a fraction of the nodes and $\bar{k}_e$ is the branching factor, we can write a mean-field approximation based on the branching process as follows:

$$I_n^{(t+1)} = \bar{k}_e[\pi_{\mathrm{nn}}I_n^{(t)} + \pi_{\mathrm{an}}I_a^{(t)}], \quad (16)$$

$$I_a^{(t+1)} = \bar{k}_e[\pi_{\mathrm{na}}I_n^{(t)} + \pi_{\mathrm{aa}}(1 - p_{\mathrm{app}})I_a^{(t)}]. \quad (17)$$

By defining $a = \pi_{nn}\bar{k}_e$, $b = \pi_{an}\bar{k}_e$, $c = \pi_{na}\bar{k}_e$, and $d = \pi_{aa}\bar{k}_e(1 - p_{app})$, the difference equations can be written as:

$$\mathbf{X}_{t+1} = \mathbf{A}\mathbf{X}_t, \tag{18}$$

where $\mathbf{X}_t = [\begin{smallmatrix} I_n^{(t)} \\ I_a^{(t)} \end{smallmatrix}]$ and $\mathbf{A} = (\begin{smallmatrix} a & b \\ c & d \end{smallmatrix})$.

The steady state $\mathbf{X}_{t+1} = \mathbf{X}_t$ is possible if all the eigenvalues λ of the transition matrix $\mathbf{A}$ (whether real or complex) have an absolute value which is less than 1;

$$\lambda_\pm = \frac{a+d}{2} \pm \sqrt{\left(\frac{a+d}{2}\right)^2 - (ad - bc)}. \tag{19}$$

Without contact tracing, there is a chance of epidemic, given the initial reproductive number is $\bar{k}_e > 1$. In the case of app adoption, the critical value of app-users π_a^c that is needed for reducing the reproductive number can be derived by setting $\lambda = 1$ which leads to:

$$\frac{1 - \pi_a(2 - \pi_{aa})}{1 - \pi_a}\left[\bar{k}_e + \bar{k}_e^2\pi_{aa}(1 - p_{app})\right]$$
$$+ \bar{k}_e\pi_{aa}(1 - p_{app}) + \frac{\bar{k}_e^2\pi_a(1 - \pi_{aa})^2}{1 - \pi_a} = 1. \tag{20}$$

When apps work perfectly, the epidemic threshold is given by:

$$\bar{k}_c = \frac{\sqrt{1 + \pi_a\pi_{aa}[4(\pi_a + \pi_{aa}) - 3(\pi_a\pi_{aa} + 2)]}}{2\pi_a(\pi_{aa} - 1)^2}$$
$$+ \frac{2\pi_a - \pi_a\pi_{aa} - 1}{2\pi_a(\pi_{aa} - 1)^2}. \tag{21}$$

For each value of π_a there is a nontrivial optimum value π_{aa}^{opt} that leads to the largest epidemic threshold in terms of the branching factor, which is

$$\pi_{aa}^{opt} = \frac{\pi_a - 2}{3\pi_a - 4}. \tag{22}$$

The critical app adoption can be also calculated as:

$$\pi_a^c = \frac{1 - \bar{k}_e}{\bar{k}_e^2(\pi_{aa} - 1)^2 + \bar{k}_e(\pi_{aa} - 2) + 1}. \tag{23}$$

In the absence of homo- or heterophily, $\pi_{aa} = \pi_a$, Eq. (20), gives the same result as of Ref. [44], such that:

$$\pi_a^c = \frac{\bar{k}_e - 1 + \sqrt{(\bar{k}_e - 1)(\bar{k}_e + 3)}}{2\bar{k}_e}. \tag{24}$$

Vazquez [97] also provides a clear way of combining different intervention strategies and shows how our specific results about application homophily are affected by other interventions.

C. Component size simulations

Next, we describe how to extract the giant component in simulated networks and how these simulation results can be

used to find the critical points of the disease spreading process. The component sizes can also be used to find the epidemic size distributions as described in Sec. II B.

1. Component extension

In each simulation run, we simulate one network structure G and distribute the apps to the nodes according to one of the models described in Sec. II C. From the original network G, we keep each link with probability $p = 1 - e^{-\beta\tau}$, which is the probability of infection going through a link without apps. We also remove all the links between two app-users with probability p_{app} and call the resulting network G_a. The components of graph G_a are the normal components.

The extended components can be reached by going through every normal component and extending it. For every app-user α in the component C, we go through the neighbors $n_\alpha = \{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ in the original network G. If α_i is an app-user and not in the component $\alpha_i \notin C$, then we add it to the component extension C' with probability p. The total set of infected nodes, if starting from a node in C, will be $C \cup C'$. As these are disjoint sets, we can compute the size as $S_C' = |C| + |C'|$ and $S_c = |C|$.

2. Susceptibility

In numerical simulations of finite-size systems, we can use the peak of a susceptibility measure to find the critical transition point. Theoretically, susceptibility [84] is a measure of fluctuation in the component sizes, which is singular at the epidemic threshold (the critical point). In network percolation studies, it is defined as the expected growth in the size of the giant component when a random link is added to the network. Therefore, susceptibility in an ordinary percolation problem can be written as:

$$\chi = \frac{\sum_{c \neq c_{max}} S_c^2 - S_{c_{max}}^2}{N - S_{c_{max}}}, \tag{25}$$

where S_c is the size of the component c, $c_{max} = \text{argmax}_c S_c$ is the largest component.

Here, we are dealing with two types of components, and as is shown in Fig. 2(d), the fraction of the sum of component sizes and network size $S_\sum/N$ can be larger than one. Susceptibility should be a monotonically decreasing function in the supercritical regime. However, plugging the extended component sizes into Eq. (25) results in a growth in the tail of susceptibility, turning it to a nonmonotonic function in the supercritical regime. Therefore, this formulation of susceptibility is not suitable in the current case since the maximum of Eq. (25) could lead to estimates of critical points that are very far from the actual one. Instead, we can use the expected growth in the extended giant component, which can be computed as:

$$\chi' = \frac{\sum_{c \neq c_{max}} S_c S_c'\left(1 - \frac{S_{c_{max}}'}{N}\right)}{N - S_{c_{max}}}, \tag{26}$$

where S_c and S_c' are the size and the extended size of the component c and $c_{max} = \text{argmax}_c S_c'$ is the largest component measured in the extended size.
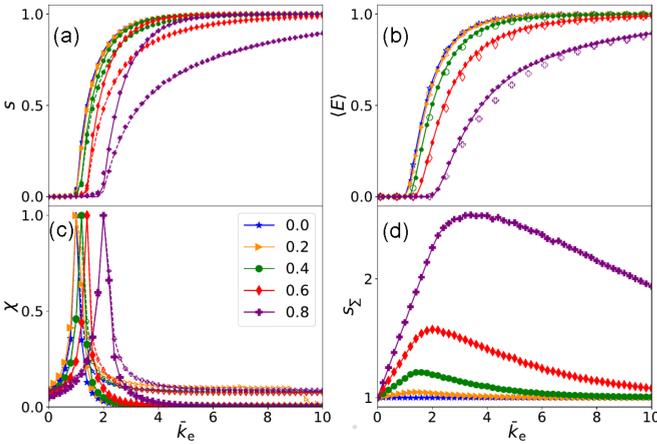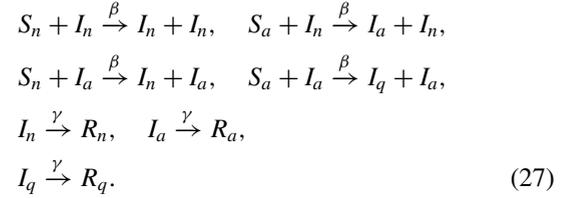
FIG. 2. Disease spreading statistics in an Erdős-Rényi network as a function of the effective connectivity $\bar{k}_e$ when there are $\pi_a N$ perfect applications ($p_{app} = 1$) that are distributed uniformly randomly. Results are normalized to the network size N and shown for $\pi_a \in [0, 0.2, 0.4, 0.6, 0.8]$ with different markers. (a) The normal component size, i.e., the epidemic probability (dashed lines and markers following them) and the extended components, i.e., the epidemic size (solid lines and markers following them). Dashed and solid lines indicate the results from theory introduced in Sec. III A by Eq. (4) and the markers are results computed from component sizes of simulated networks as described in Sec. III C. (b) The expected epidemic size as given by Eq. (1) computed with theoretical results introduced in Sec. III A (solid lines), simulated component sizes introduced in Sec. III C (filled markers), and explicit SIR simulations introduced in Sec. III D (empty markers). (c) Susceptibility of the normal giant component χ (dots) and the extended component χ' (solid lines) as defined in Eqs. (25) and (26). Since susceptibility is a divergent quantity at the epidemic threshold, as explained in Sec. III C 2, it is a good proxy for finding the critical point. Notice that peaks are at the same positions for both curves, normal and extended components. (d) The fraction of sum of component sizes and network size S_Σ / N.

D. Explicit compartment model simulations

Finally, we will perform explicit simulations of the spreading processes to confirm the theoretical results we arrived at via the approximations we presented above. The effect of tracking applications can be integrated into compartment model simulation by introducing separate susceptible and infected compartments for people with and without the app. The interactions between people with no app installed is similar to those of the normal SIR process, namely, susceptible individuals with no app (S_n) can become infected (I_n) by being in contact with infected people that either do not have the app installed (I_n) or have it installed (I_a). However, if a susceptible individual with the app (S_a) comes into contact with an infected individual with app (I_a), they will become infected but they will also receive infection notification from the app which means they will be quarantined (I_q). Quarantined individuals cannot infect anyone else. Eventually, all the infected individuals will move to the recovered compartment after a constant predetermined amount of time ($1/\gamma$) has passed from the beginning of their infection. The recovered compartment

is divided into three compartments R_n, R_a, and R_q to track which infected compartment the node is originating from.

The set of all reactions can be written as follows:

$$
\begin{aligned}
S_n + I_n &\xrightarrow{\beta} I_n + I_n, \quad S_a + I_n \xrightarrow{\beta} I_a + I_n, \\
S_n + I_a &\xrightarrow{\beta} I_n + I_a, \quad S_a + I_a \xrightarrow{\beta} I_q + I_a, \\
I_n &\xrightarrow{\gamma} R_n, \quad I_a \xrightarrow{\gamma} R_a, \\
I_q &\xrightarrow{\gamma} R_q.
\end{aligned}
\tag{27}
$$

Note that while edge reactions are governed by Poisson processes happening at a constant rate β, unlike most common SIR models, node reactions are governed by constant cutoff time $1/\gamma$ and happen exactly $1/\gamma$ units of time after the infection of the node.

As interactions in the simulation are bound to take place over edges of a static network, with nodes belonging to each of the compartments, as shown in Sec. IV, the results are similar to a component size simulation (which are described in Sec. III C) on a network with effective connectivity of $\bar{k}_e = \langle k \rangle (1 - e^{-\beta/\gamma})$. As only the ratio between β and γ plays as a parameter in the model, we set the value of γ to 1.

In each simulation, starting from a single infected node and running the simulation in discrete time steps of 10^{-4} units until no further reaction is possible, the final number of nodes that end up in R_q, R_a, and R_n determine total size of infection corresponding to the extended component size S' of the component that the initial seed node belongs to. The final combined size of the R_n and R_a component, however, represents the size of the component S_n that the seed node (index case) would belong to, had we removed app-app links. By adding I_a and I_q compartments, as compared to normal SIR processes, and linking them to the state of the source of infection and the internal state of each node, we include information about the history of the spreading agent more than one step back in the simulation of the spreading process.

IV. NUMERICAL RESULTS

We will next illustrate using the theory and simulation introduced in Sec. III how the various parameters affect the epidemic sizes and epidemic probabilities. The simulation studies are done in networks of 10^4 nodes and averaged over 10 realizations. We use two network topologies: homogeneous networks (Erdős-Rényi networks) with expected degree $\langle k \rangle = 10$ and random networks with expected degree sequence driven from power-law degree distribution $p(k) \propto k^{-3}$, with a minimum degree cutoff adjusted such that the average degree is set to 10 [94].

A. Differences in normal and extended components

The difference between the epidemic probability (normal component size) and the epidemic size (extended component size), as given by Eqs. (4) and (5), is a phenomenon specific to epidemics in the presence of app-adaptors. Breaking the equivalence of these two measures can have practical consequences, as illustrated in Fig. 2(a). The difference between these two grows with the fraction of app-users π_a. For example, when $\pi_a = 0.8$ and the epidemic probability
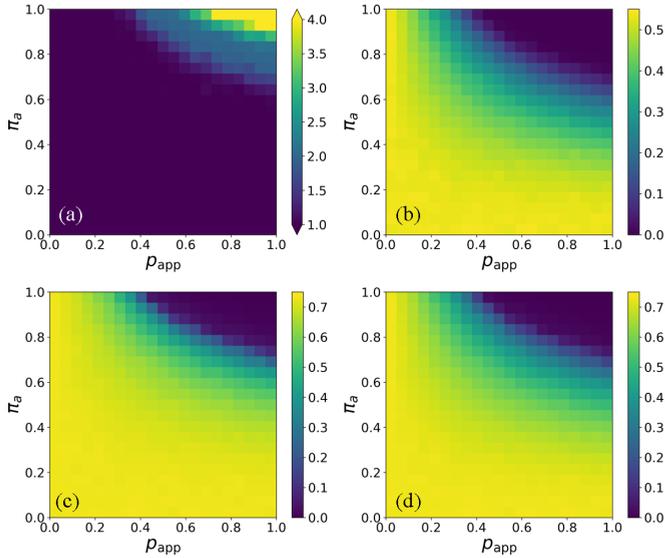
FIG. 3. The effect of quarantine failures as described in Sec. IV B in homogeneous networks when app adoption is done uniformly randomly. Results are from percolation simulations. (a) The epidemic threshold as a function of quarantine probability p_{app} and app adoption rate π_a. All threshold values larger than 4 are shown with the same color. By setting the effective connectivity of the network to $\bar{k}_e = 1.8$ (b) the expected epidemic size, (c) the extended giant component size, and (d) the normal giant component size are shown as a function of p_{app} and π_a. Note that $\bar{k}_e = 1.8$ is chosen as an illustrative example of a parameter region with interesting behavior in the various component sizes: It is large enough such that without any intervention, there is a wide epidemic spreading, but small enough such that the spread can be controlled without extreme measures.

(the normal component size) is $s_{max} \approx 0.5$, the epidemic size (the extended component size) reaches $s_{max} \approx 0.8$. This is also reflected in the expected epidemic sizes [see Fig. 2(b) and Eq. (1)]. Despite the two component definitions differing from each other, they still display the transition at the same point and this point can be measured numerically using the susceptibilities defined in Eqs. (25) and (26) [see Fig. 2(c)].

The extended component size is not a conserved quantity like the normal component size in the sense that the sum of component sizes S_Σ would always sum to the number of nodes N. Instead, the sum of component sizes can be significantly larger than the number of nodes [see Fig. 2(d)] and the maximum value it can reach grows with the number of application users π_a. The deviation from $S_\Sigma/N = 1$ reaches its maximum with disease parameters higher than the threshold values, but when the disease reaches a large enough population, the fraction S_Σ/N starts to decay, reaching $S_\Sigma/N = 1$ when everybody belongs to the normal giant component.

B. Quarantine failures

The assumption in Sec. IV A is that (i) apps work perfectly and (ii) an app-user always self-isolates before having a chance to spread the infection, meaning that there are no quarantine failures, $p_{app} = 1$. It is of practical significance to investigate the effects of quarantine failures [45] on the epidemic threshold and epidemic size. Figure 3 shows that
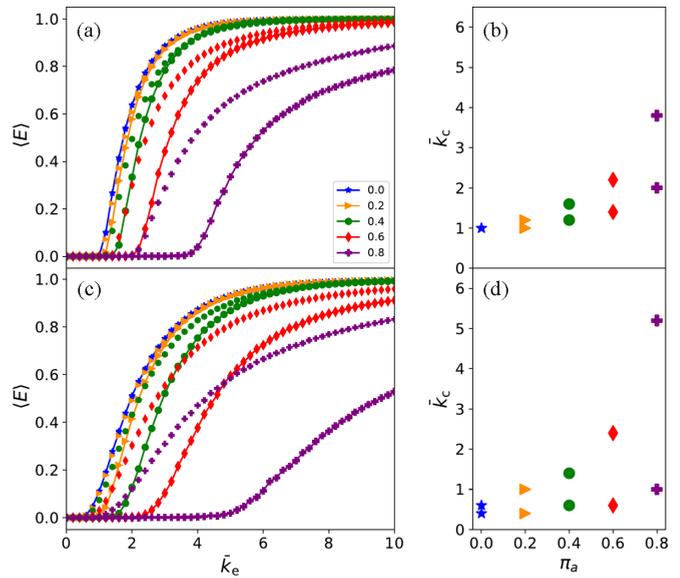

FIG. 4. Expected epidemic size $\langle E \rangle$ and epidemic threshold $\bar{k}_c$ for two network topologies with different strategies; $\langle E \rangle$ as a function of effective connectivity $\bar{k}_e$ for (a) homogeneous networks with Poisson degree distribution and for (c) heterogeneous networks with a power-law degree distribution $P(k) \propto k^{-3}$. Results are shown for different values of π_a using different markers: 0 (stars), 0.2 (triangles), 0.4 (disks), 0.6 (diamonds), and 0.8 (crosses). The solid lines with markers indicate the high-degree targeting strategy, while single markers indicate the random app adoption. Epidemic threshold $\bar{k}_c$ as a function of app-adoption rate π_a (such that the upper markers represent the high-degree targeting strategy) for (b) homogeneous networks and for (d) heterogeneous networks. Differences between the threshold values in the presence of homophily are explained in Figs. 5(b) and 5(d).

in the absence of major quarantine failures, epidemic tracing and mitigation with apps can still be a valid strategy if the app adoption level in a society is high enough. The effect of app adoption rate π_a is more important than the rate at which apps function, but both need to be relatively high in order for the apps to have a significant impact.

Even if we are above the epidemic threshold, the apps can be useful. Especially when the application adoption π_a is high, the quarantines can be very unreliable and the outbreak size [Figs. 3(b) and 3(c)] and epidemic probability [Fig. 3(d)] both remain small. Again, overall, app adoption and quarantine reliability are essential, with the app adoption rate being more important.

C. Degree heterogeneity and high-degree app targeting

Real networks are degree-heterogeneous and this heterogeneity has a strong effect on the final outbreak size and the epidemic threshold. Figure 4 shows the expected epidemic sizes with two different strategies in app adoption, random, and high-degree targeting, for different fractions of app-users π_a in the network. In homogeneous networks, Fig. 4(a), contact tracing decreases the expected epidemic size and pushes the epidemic threshold forward. These effects can be further amplified by shifting to the high-degree targeting in
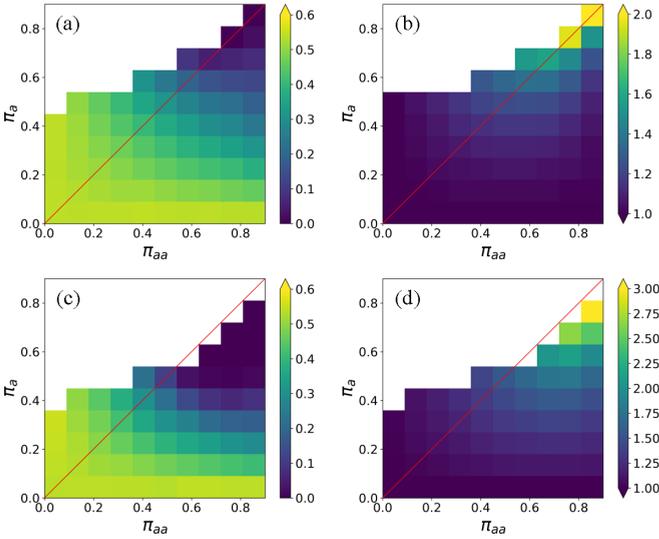
FIG. 5. The effect of homophily or heterophily in app adoption in homogeneous networks as described in Sec. IV D. Homophily (heterophily) region is below (above) the diagonal $\pi_a = \pi_{aa}$. Expected epidemic size at $\bar{k}_e = 1.8$ for (a) random app adoption and for (c) high-degree targeting strategy. The epidemic threshold for (b) random app adoption and for (d) the high-degree targeting strategy. Thresholds are from theoretical results given by Eq. (21) and expected epidemic sizes are from percolation simulations. The empty white region is the spectrum that having such a homo- or heterophilic population is impossible.

app adoption. With 80% of app-users, the epidemic threshold can move from $\bar{k}_e = 1$ to $\bar{k}_e = 4$, which means at that point expected epidemic size is zero, while without contact tracing it would be almost 1. Note that in homogeneous networks, the effective average degree of the contact network $\bar{k}_e$, has good correspondence to the reproduction number of the infection.

In networks with degree-heterogeneity, the epidemic threshold vanishes in normal SIR processes. This effect holds in contact-traced epidemics if we distribute the apps uniformly randomly. However, from Fig. 4(b) it is clear that contact tracing can significantly reduce the expected epidemic size even when the apps are randomly distributed and the epidemic threshold remains unchanged. With the high-degree targeting strategy, it is possible to move the epidemic threshold. Comparing the expected epidemic size at different values of $\bar{k}_e < 3$ shows that in real-world situations, app adoption of superspreaders is of significant importance. Since hubs become the app-users, this strategy has drastic effects on the size and threshold of the epidemic, such that the threshold gets pushed from somewhere near zero to a value $\bar{k}_e > 5$ with the app adoption rate $\pi_a = 0.8$. Therefore, the reproduction number can be much more controlled in the high-degree targeting strategy.

D. The effect of homophily and heterophily

In previous sections, there was an assumption that app-users are distributed with random mixing patterns; the fact that one of the connections of a node is an app-user has no effect on the probability of that node being
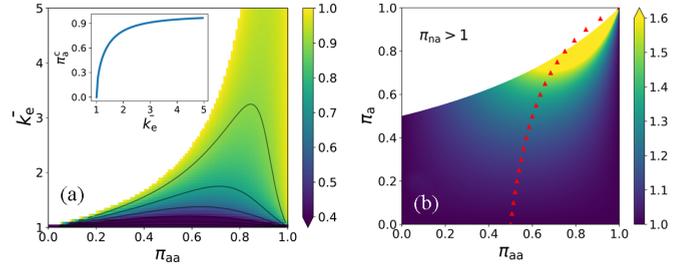

FIG. 6. Existence of optimum value for homophily based on branching process approximation as described in Sec. III B. (a) The critical value of app-users π_a^c that are needed for reducing the reproductive number as a function of effective connectivity and homophily probability π_{aa}. The value of π_a^c remains the same within each black curve. The inset is the graph of π_a^c as a function of $\bar{k}_e$ in the absence of homophily $\pi_{aa} = \pi_a$ given by Eq. (24). (b) The epidemic threshold $\bar{k}_e$ as a function of π_{aa} and π_a. The red symbols show the π_{aa}^{opt} for each π_a which is given by to Eq. (22). The pattern here is consistent with another approximation shown in Fig. 5(b), while epidemic threshold values are slightly different due to different levels of approximations. Note that here we display the epidemic threshold for all values of π_{aa} and π_a such that $0 \leqslant \pi_{na} \leqslant 1$ so the networks with some of these parameters can be created in practice [95].

an app adopter. Next, we explore how homophily or heterophily affects epidemics based on app usage using the MN model. A Swiss experiment has reported that while a small fraction of $\pi_a = 0.2$ of people have used the app, the inside connections between them was high enough such that $\pi_{aa} = 0.7$ [40].

Figure 5 illustrates that increasing heterophily leads to a lower epidemic threshold and larger epidemic size for a fixed $\bar{k}_e$. Increasing homophily from random mixing is initially preferable, but the optimum lies between random mixing and full homophily. For the expected epidemic size, strong heterophily is especially detrimental [see Fig. 5(a) for the homogeneous network and with random app adoption and in Fig. 5(c) for high-degree targeting strategy]. The optimum value for heterophily or homophily is evident for the epidemic thresholds in Figs. 5(b) and 5(d), respectively, for the random and high-degree targeting strategies. Figure 6(b) gives a more clear picture of existence of an optimum value for the epidemic threshold in the case of homophily. According to Eq. (21), for each fraction of app-users π_a in the network, the epidemic threshold $\bar{k}_c(\pi_a, \pi_{aa})$ can be maximized by controlling the homophily in app adoption π_{aa}. The pattern in the Fig. 6(b) is very similar to the convex pattern in Fig. 5(b), even though they are calculated using different approximations and approaches (see Secs. III A and III B).

Another view on the effect of homophily and heterophily is given by finding the critical fraction app-users π_a^c that is needed to go beyond the epidemic threshold as a function of (π_{aa} and $\bar{k}_e$). Figure 6(a) depicts this relationship based on Eq. (23) and shows that π_a^c is not monotonic function of π_{aa} but there is an optimal value of π_{aa} giving the lowest fraction apps that are needed to stop the epidemic. Note that in a network without homophily or heterophily π_a^c increases monotonically as the function of the effective connectivity $\bar{k}_e$ [see the inset of Fig. 6(a)].
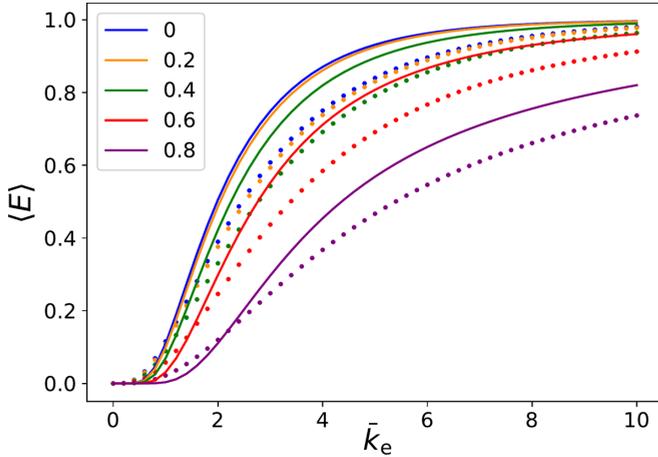
FIG. 7. The expected epidemic size computed with theoretical results introduced in Sec. III A for heterogeneous networks with degree distribution $P(k) \propto k^{-3}$ (solid lines) compared with ones with $P(k) \propto k^{-2.5}$ (dotted lines) as a function of the effective connectivity $\bar{k}_e$ when apps are distributed uniformly randomly. Results are normalised to the network size N and shown for $\pi_a \in [0, 0.2, 0.4, 0.6, 0.8]$ with different colors. Note that by lowering the exponent, epidemic thresholds get closer to zero and the expected epidemic sizes decrease since there more low-degree nodes in the network. Therefore, by lowering the exponent, while we can add more degree heterogeneity in the network, the physics of the phenomena does not change.

V. DISCUSSIONS

In this article, we have developed two flexible analytic approximations to SIR epidemics in the presence of contact tracing apps. First, we use a branching process to derive explicit analytical solutions for the epidemic thresholds. Second, we expand the framework of using self-consistent equations to analyze digital contact tracing [44], which is an alternative to other approaches [71]. Contrary to the conventional SIR spreading, a full picture of the late-state epidemics in the presence of digital contact tracing is not given by a single observable (the component size), but one also needs two variables (normal and extended component sizes). These correspond to the probability of the epidemic and the epidemic size, which are equivalent in the SIR process. Here we see that the two quantities can be significantly different if the number of application users is high.

Our numerical results illustrate that the effects of digital contact tracing can be very sensitive to the network structure, how applications are distributed among the population, and how well the tracing works. Realistic estimates of the effects of digital contact tracing can only be achieved if one can choose correct parameter ranges in a high-dimensional parameter space. In this study, we had six of such parameters: the shape of the degree distribution, average degree, amount of heterophily or homophily, application prevalence, quarantine probability and targeting strategy. While we were able to establish and confirm basic laws governing individual parameters and some combinations of parameters, exploring such a parameter space fully for possible compound effects is out of the reach in simulations. However, these effects can be largely revealed by inspecting the analytic equations we derived.
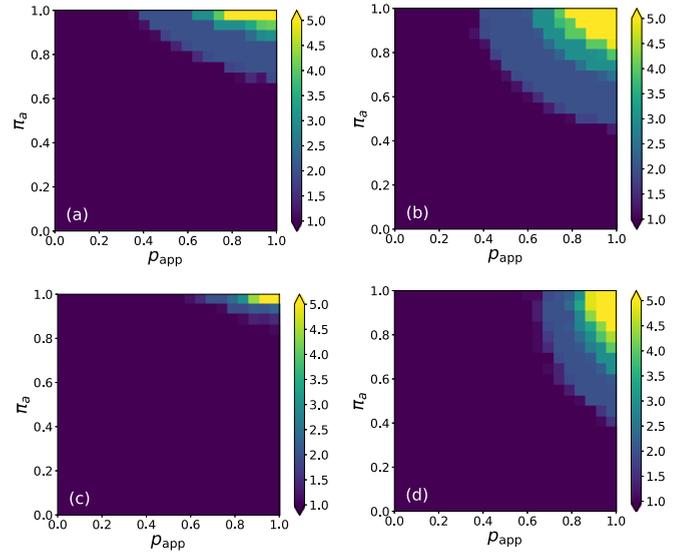

FIG. 8. The epidemic threshold as a function of quarantine probability p_{app} and app adoption rate π_a. The effect of quarantine failures in homogeneous networks with (a) random app adoption (b) and high-degree targeting strategy. Also, for heterogeneous networks with a power-law degree distribution with (c) random app adoption (d) and high-degree targeting strategy. All threshold values larger than 5 are shown with the same color.

There are several open questions for which this study and other studies only hint at the results. There are types of network structures we ignore here. For example, the heterophily
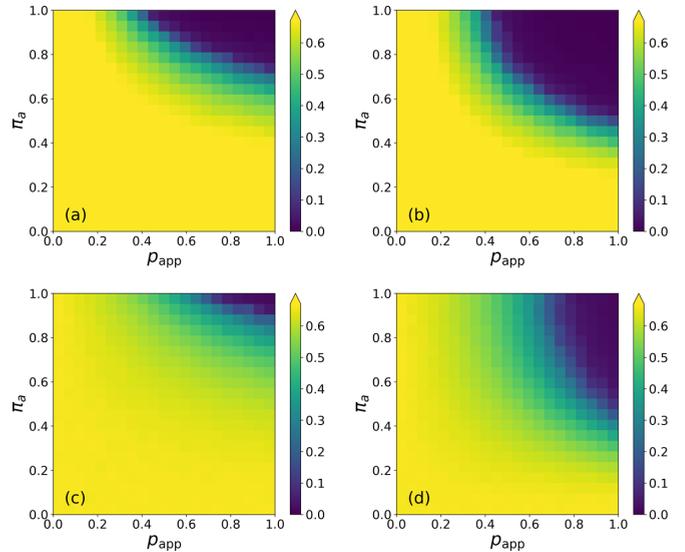

FIG. 9. Expected epidemic size in the case of quarantine failures. Expected epidemic size at $\bar{k}_e = 1.8$ for homogeneous networks with (a) random app adoption (b) and high-degree targeting strategy. Also, for heterogeneous networks with a power-law degree distribution with (c) random app adoption (d) and high-degree targeting strategy. In (b) and (d) the pattern is different due to the effects of hubs. When doing a high-degree targeting strategy, quarantine failures are more significant since the infected ones are highly influential on the spreading dynamics.
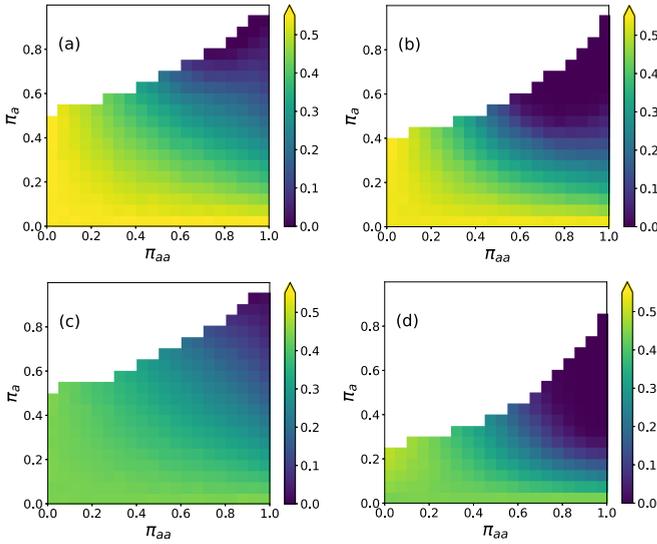
FIG. 10. The effect of homophily or heterophily in app adoption on the expected epidemic size. Expected epidemic size at $\bar{k}_e = 1.8$ from percolation simulations for homogeneous networks with (a) random app adoption (b) and high-degree targeting strategy. Also, for heterogeneous networks with a power-law degree distribution with (c) random app adoption and (d) high-degree targeting strategy. The empty white region is the spectrum that having such a homo- or heterophilic population is impossible.
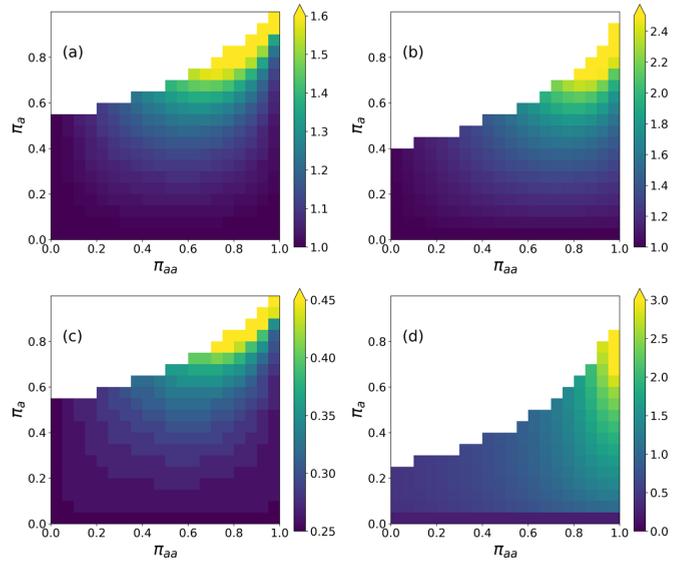

FIG. 11. The effect of homophily or heterophily in app adoption on the epidemic threshold and optimum pattern for homophily. Epidemic thresholds for homogeneous networks with (a) random app adoption (b) and high-degree targeting strategy. Also, for heterogeneous networks with a power-law degree distribution with (c) random app adoption and (d) high-degree targeting strategy. The empty white region is the spectrum that having such a homo- or heterophilic population is impossible.

and homophily could be constructed in the network in slightly different ways. For example, a case study using a realistic agent-based model [69] has recently considered, among many other modeling choices aimed at precise calibration on the French population, the contributions of individuals of different ages. One could also develop a more realistic version of our stylized model to systematically analyze the effects of homophily caused by an age-based contact structure and different scenarios of app adoption within that structure. The age-based approach would also allow one to estimate the benefits of applications relative to the risk groups in this model.

Overall the problem of digital contact tracing offers not only a practical problem to solve but also an interesting theoretical puzzle because it introduces memory to the epidemic process. This memory is limited to one step within the tracing model we use here, but one could also use multistep tracing, where also the second neighbors of infected nodes are quarantined in the case that the first neighbors have already passed on the infection. Further, here we ignore effects such as quarantines that do not directly stop the infection from one application user to another from spreading further. However, in the case of a strong group structure in the network, there could be situations where a nonapplication user A infects application user B, who alerts another application user C, who actually gets infected by A and stops the spreading because of the quarantine. Analyzing such more complicated phenomena can provide challenges for network scientists for years to come.

ACKNOWLEDGMENTS

APPENDIX

The heterogeneity in the number of contacts could also be modeled with other distributions, for example, the negative binomial distribution. This would have the advantage of having a nodivergent second moment supported by empirical evidence. However, we aimed to illustrate the effects of degree heterogeneity and not perform a systematic analysis. We already have many different random network models and combinations of parameters related to the app distribution, how well it works, and disease parameters. The equations we give make it possible for one to do such analysis if needed. Therefore we limited our main discussion to the differences observed in a power-law network with exponent -3 compared to the results for homogeneous networks. However, to satisfy the curiosity of the reader interested in extreme heterogeneity, we have now added Fig. 7 showing the expected epidemic size for exponent -2.5.

About quarantine failures, as it was shown in Figs. 3, 8 and 9 also show that contact tracing can yield very good results in terms of reducing the epidemic threshold and expected epidemic size if everything goes right at least for 50% and half of the people use the apps. This effect is more prominent if we go for the high-degree targeting strategy, especially in heterogeneous networks, as shown in Figs. 8(d) and 9(d). Figures 10 and 11 show that there is an optimum value for homophily in app adoption as it was shown in Fig. 5 and Fig. 6. The only exception is when we follow a high-degree targeting strategy in heterogeneous networks. In this case, we can see the hub effect on the epidemic threshold and size.

[1] N. Perra, Non-pharmaceutical interventions during the covid-19 pandemic: A review, Phys. Rep. **913**, 1 (2021).

[2] Y. Cheng, N. Ma, C. Witt, S. Rapp, P. S. Wild, M. O. Andreae, U. Pöschl, and H. Su, Face masks effectively limit the probability of sars-cov-2 transmission, Science **372**, 1439 (2021).

[3] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, Princeton, NJ, 2011).

[4] W. H. Foege, J. D. Millar, and J. M. Lane, Selective epidemiologic control in smallpox eradication, Am. J. Epidemiol. **94**, 311 (1971).

[5] K. C. Swanson, C. Altare, C. S. Wesseh, T. Nyenswah, T. Ahmed, N. Eyal, E. L. Hamblion, J. Lessler, D. H. Peters, and M. Altmann, Contact tracing performance during the ebola epidemic in liberia, 2014-2015, PLoS Negl. Trop. Dis. **12**, e0006762 (2018).

[6] G. W. Rutherford and J. M. Woo, Contact tracing and the control of human immunodeficiency virus infection, J. Am. Med. Assoc. **259**, 3609 (1988).

[7] G. J. Fox, S. E. Barry, W. J. Britton, and G. B. Marks, Contact investigation for tuberculosis: a systematic review and meta-analysis, Eur. Respir. J. **41**, 140 (2013).

[8] A. Aleta, D. Martin-Corral, A. P. y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini Jr, S. Merler *et al.*, Modelling the impact of testing, contact tracing and household quarantine on second waves of covid-19, Nat. Hum. Behav. **4**, 964 (2020).

[9] P. G. Walker, C. Whittaker, O. J. Watson, M. Baguelin, P. Winskill, A. Hamlet, B. A. Djafaara, Z. Cucunubá, D. O. Mesa, W. Green *et al.*, The impact of covid-19 and strategies for mitigation and suppression in low-and middle-income countries, Science **369**, 413 (2020).

[10] S. M. Kissler, C. Tedijanto, E. Goldstein, Y. H. Grad, and M. Lipsitch, Projecting the transmission dynamics of sars-cov-2 through the postpandemic period, Science **368**, 860 (2020).

[11] J. Bell, G. Bianconi, D. Butler, J. Crowcroft, P. C. Davies, C. Hicks, H. Kim, I. Z. Kiss, F. Di Lauro, C. Maple *et al.*, Beyond covid-19: Network science and sustainable exit strategies, J. Phys. Complex. **2**, 021001 (2021).

[12] M. Gilbert, M. Dewatripont, E. Muraille, J.-P. Platteau, and M. Goldman, Preparing for a responsible lockdown exit strategy, Nat. Med. **26**, 643 (2020).

[13] N. E. Kogan, L. Clemente, P. Liautaud, J. Kaashoek, N. B. Link, A. T. Nguyen, F. S. Lu, P. Huybers, B. Resch, C. Havas, A. Petutschnig, J. Davis, M. Chinazzi, B. Mustafa, W. P. Hanage, A. Vespignani, and M. Santillana, An early warning approach to monitor covid-19 activity with multiple digital traces in near real time, Sci. Adv. **7**, eabd6989 (2021).

[14] S. Kojaku, L. Hébert-Dufresne, E. Mones, S. Lehmann, and Y.-Y. Ahn, The effectiveness of backward contact tracing in networks, Nat. Phys. **17**, 652 (2021).

[15] M. Salathe, L. Bengtsson, T. J. Bodnar, D. D. Brewer, J. S. Brownstein, C. Buckee, E. M. Campbell, C. Cattuto, S. Khandelwal, P. L. Mabry *et al.*, Digital epidemiology, PLoS Comput. Biol. **8**, e1002616 (2012).

[16] A. Natarajan, H.-W. Su, and C. Heneghan, Assessment of physiological signs associated with covid-19 measured using wearable devices, npj Digit. Med. **3**, 156 (2020).

[17] D. R. Seshadri, E. V. Davies, E. R. Harlow, J. J. Hsu, S. C. Knighton, T. A. Walker, J. E. Voos, and C. K. Drummond, Wearable sensors for covid-19: A call to action to harness our digital infrastructure for remote patient monitoring and virtual assessments, Front. Digit. Health **2**, 8 (2020).

[18] D. P. Oran and E. J. Topol, Prevalence of asymptomatic sars-cov-2 infection: A narrative review, Ann. Intern. Med. **173**, 362 (2020).

[19] E. Morales-Narváez and C. Dincer, The impact of biosensing in a pandemic outbreak: Covid-19, Biosens. Bioelectron. **163**, 112274 (2020).

[20] G. Quer, J. M. Radin, M. Gadaleta, K. Baca-Motes, L. Ariniello, E. Ramos, V. Kheterpal, E. J. Topol, and S. R. Steinhubl, Wearable sensor data and self-reported symptoms for covid-19 detection, Nat. Med. **27**, 73 (2021).

[21] T. Mishra, M. Wang, A. A. Metwally, G. K. Bogu, A. W. Brooks, A. Bahmani, A. Alavi, A. Celli, E. Higgs, O. Dagan-Rosenfeld *et al.*, Pre-symptomatic detection of covid-19 from smartwatch data, Nat. Biomed. Eng. **4**, 1208 (2020).

[22] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van den Broeck, C. Régis, B. Lina *et al.*, High-resolution measurements of face-to-face contact patterns in a primary school, PLoS ONE **6**, e23176 (2011).

[23] A. Ganguli, A. Mostafa, J. Berger, M. Y. Aydin, F. Sun, S. A. S. de Ramirez, E. Valera, B. T. Cunningham, W. P. King, and R. Bashir, Rapid isothermal amplification and portable detection system for sars-cov-2, Proc. Natl. Acad. Sci. USA **117**, 22727 (2020).

[24] P. Fozouni *et al.*, Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy, Cell **184**, 323 (2021).

[25] C. Menni *et al.*, Real-time tracking of self-reported symptoms to predict potential COVID-19, Nat. Med. **26**, 1037 (2020).

[26] T. Varsavsky, M. S. Graham, L. S. Canas, S. Ganesh, J. C. Pujol, C. H. Sudre, B. Murray, M. Modat, M. J. Cardoso, C. M. Astley *et al.*, Detecting covid-19 infection hotspots in england using large-scale self-reported data from a mobile application: A prospective, observational study, Lancet Publ. Health **6**, e21 (2021).

[27] A. Ganguli, A. Ornob, H. Yu, G. Damhorst, W. Chen, F. Sun, A. Bhuiya, B. Cunningham, and R. Bashir, Hands-free smartphone-based diagnostics for simultaneous detection of zika, chikungunya, and dengue at point-of-care, Biomed Microdev. **19**, 73 (2017).

[28] L. B. Sousa, S. R. Fricker, S. S. Doherty, C. E. Webb, K. L. Baldock, and C. R. Williams, Citizen science and smartphone e-entomology enables low-cost upscaling of mosquito surveillance, Sci. Total Environ. **704**, 135349 (2020).

[29] N. G. Becker, K. Glass, Z. Li, and G. K. Aldis, Controlling emerging infectious diseases like sars, Math. Biosci. **193**, 205 (2005).

[30] M. Eichner, Case isolation and contact tracing can prevent the spread of smallpox, Am. J. Epidemiol. **158**, 118 (2003).

[31] K. T. Eames and M. J. Keeling, Contact tracing and disease control, Proc. R. Soc. Lond. B **270**, 2565 (2003).

[32] I. Z. Kiss, D. M. Green, and R. R. Kao, Infectious disease control using contact tracing in random and scale-free networks, J. R. Soc. Interface **3**, 55 (2006).

[33] R. M. HOWELL, W. J. Kassler, and A. Haddix, Partner notification to prevent pelvic inflammatory disease in women: Cost-effectiveness of two strategies, Sex. Transmit. Dis. **24**, 287 (1997).

[34] C. Fraser, S. Riley, R. M. Anderson, and N. M. Ferguson, Factors that make an infectious disease outbreak controllable, Proc. Natl. Acad. Sci. USA **101**, 6146 (2004).

[35] M. Kretzschmar, Y. T. van Duynhoven, and A. J. Severijnen, Modeling prevention strategies for gonorrhea and chlamydia using stochastic network simulations, Am. J. Epidemiol. **144**, 306 (1996).

[36] D. Klinkenberg, C. Fraser, and H. Heesterbeek, The effectiveness of contact tracing in emerging epidemics, PLoS ONE **1**, e12 (2006).

[37] R. Huerta and L. S. Tsimring, Contact tracing and epidemics control in social networks, Phys. Rev. E **66**, 056115 (2002).

[38] P. Rodríguez, S. Graña, E. E. Alvarez-León, M. Battaglini, F. J. Darias, M. A. Hernán, R. López, P. Llaneza, M. C. Martín, O. Ramirez-Rubio *et al.*, A population-based controlled experiment assessing the epidemiological impact of digital contact tracing, Nat. Commun. **12**, 587 (2021).

[39] I. Braithwaite, T. Callender, M. Bullock, and R. W. Aldridge, Automated and partly automated contact tracing: A systematic review to inform the control of covid-19, Lancet Digit. Health (2020).

[40] M. Salathé, C. L. Althaus, N. Anderegg, D. Antonioli, T. Ballouz, E. Bugnion, S. Capkun, D. Jackson, S.-I. Kim, J. Larus *et al.*, Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland, Swiss Med. Weekly **150**, w20457 (2020).

[41] J. O'Connell and D. T. O'Keeffe, Contact tracing for covid-19—A digital inoculation against future pandemics, New Engl. J. Med. (2021).

[42] G. Cencetti, G. Santin, A. Longa, E. Pigani, A. Barrat, C. Cattuto, S. Lehmann, M. Salathe, and B. Lepri, Digital proximity tracing on empirical contact networks for pandemic control, Nat. Commun. **12**, 1655 (2021).

[43] H. L. Hambridge, R. Kahn, and J.-P. Onnela, Examining sars-cov-2 interventions in residential colleges using an empirical network, medRxiv 10.1101/2021.03.09.21253198 (2021).

[44] A. Barrat, C. Cattuto, M. Kivelä, S. Lehmann, and J. Saramäki, Effect of manual and digital contact tracing on covid-19 outbreaks: A study on empirical contact data, J. R. Soc. Interface **18**, rsif.2020.1000 (2021).

[45] C. Wymant, L. Ferretti, D. Tsallis, M. Charalambides, L. Abeler-Dörner, D. Bonsall, R. Hinch, M. Kendall, L. Milsom, M. Ayres *et al.*, The epidemiological impact of the nhs covid-19 app, Alan Turing Institute: London, UK (2021).

[46] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, UK, 2007).

[47] S. N. Dorogovtsev and J. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford, 2013).

[48] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, Rev. Mod. Phys. **74**, 47 (2002).

[49] R. Pastor-Satorras and A. Vespignani, Epidemic Spreading in Scale-Free Networks, Phys. Rev. Lett. **86**, 3200 (2001).

[50] M. Newman, *Networks* (Oxford University Press, Oxford, 2018).

[51] L. Hébert-Dufresne, B. M. Althouse, S. V. Scarpino, and A. Allard, Beyond r 0: Heterogeneity in secondary infections and probabilistic epidemic forecasting, J. R. Soc. Interface **17**, 20200393 (2020).

[52] M. Boguná, R. Pastor-Satorras, and A. Vespignani, Absence of Epidemic Threshold in Scale-Free Networks with Degree Correlations, Phys. Rev. Lett. **90**, 028701 (2003).

[53] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, Velocity and Hierarchical Spread of Epidemic Outbreaks in Scale-Free Networks, Phys. Rev. Lett. **92**, 178701 (2004).

[54] R. Pastor-Satorras and A. Vespignani, Epidemic dynamics in finite size scale-free networks, Phys. Rev. E **65**, 035108(R) (2002).

[55] S. Dhara, R. van der Hofstad, and J. S. van Leeuwaarden, Critical percolation on scale-free random graphs: New universality class for the configuration model, Commun. Math. Phys. **382**, 123 (2021).

[56] S. Clémençon, H. De Arazoza, F. Rossi, and V. C. Tran, A statistical network analysis of the hiv/aids epidemics in cuba, Soc. Netw. Anal. Min. **5**, 1 (2015).

[57] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz, Superspreading and the effect of individual variation on disease emergence, Nature (Lond.) **438**, 355 (2005).

[58] K. Kupferschmidt, Why do some covid-19 patients infect many others, whereas most don't spread the virus at all, Science **10**, 1464 (2020).

[59] S. L. Feld, Why your friends have more friends than you do, Am. J. Sociol. **96**, 1464 (1991).

[60] M. E. J. Newman, Threshold Effects for Two Pathogens Spreading on a Network, Phys. Rev. Lett. **95**, 108701 (2005).

[61] Y. Liu, R. M. Eggo, and A. J. Kucharski, Secondary attack rate and superspreading events for sars-cov-2, The Lancet **395**, e47 (2020).

[62] Z. Shen, F. Ning, W. Zhou, X. He, C. Lin, D. P. Chin, Z. Zhu, and A. Schuchat, Superspreading sars events, beijing, 2003, Emerg. Infect. Dis. **10**, 256 (2004).

[63] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser, Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing, Science **368**, eabb6936 (2020).

[64] P. Sapiezynski, J. Pruessing, and V. Sekara, The fallibility of contact-tracing apps, arXiv:2005.11297 (2020).

[65] M. McPherson, L. Smith-Lovin, and J. M. Cook, Birds of a feather: Homophily in social networks, Annu. Rev. Sociol. **27**, 415 (2001).

[66] S. Funk, M. Salathé, and V. A. Jansen, Modelling the influence of human behaviour on the spread of infectious diseases: a review, J. R. Soc. Interface **7**, 1247 (2010).

[67] D. Centola, An experimental study of homophily in the adoption of health behavior, Science **334**, 1269 (2011).

[68] V. von Wyl, M. Höglinger, C. Sieber, M. Kaufmann, A. Moser, M. Serra-Burriel, T. Ballouz, D. Menges, A. Frei, and M. A. Puhan, Drivers of acceptance of covid-19 proximity tracing apps in switzerland: Panel survey analysis, JMIR Publ. Health Surveill. **7**, e25701 (2021).

[69] J. A. M. López, B. A. García, P. Bentkowski, L. Bioglio, F. Pinotti, P.-Y. Boëlle, A. Barrat, V. Colizza, and C. Poletto, Anatomy of digital contact tracing: Role of age,

transmission setting, adoption and case detection, Sci. Adv. **7**, eabd8750 (2021).

[70] S. Munzert, P. Selb, A. Gohdes, L. F. Stoetzer, and W. Lowe, Tracking and promoting the usage of a covid-19 contact tracing app, Nat. Hum. Behav. **5**, 247 (2021).

[71] G. Bianconi, H. Sun, G. Rapisardi, and A. Arenas, Message-passing approach to epidemic tracing and mitigation with apps, Phys. Rev. Res. **3**, L012014 (2021).

[72] I. Kryven and C. Stegehuis, Contact tracing in configuration models, J. Phys. Complex. **2**, 025004 (2021).

[73] A. Allard, C. Moore, S. V. Scarpino, B. M. Althouse, and L. Hébert-Dufresne, The role of directionality, heterogeneity and correlations in epidemic risk and spread, arXiv:2005.11283 (2020).

[74] R. B. Simões, I. Amaral, and S. Santos, Tracking the outbreak and far beyond: How are public authorities using mobile apps to control covid-19 pandemic, Multidisc. Perspect. Commun. Pand. Context **1**, 166 (2021).

[75] M. Abueg, R. Hinch, N. Wu, L. Liu, W. J. Probert, A. Wu, P. Eastham, Y. Shafi, M. Rosencrantz, M. Dikovsky *et al.*, Modeling the effect of exposure notification and non-pharmaceutical interventions on COVID-19 transmission in Washington state, NPJ Digital Med. **4**, 1 (2021).

[76] A. Bassolas, A. Santoro, S. Sousa, S. Rognone, and V. Nicosia, Optimising the mitigation of epidemic spreading through targeted adoption of contact tracing apps, arXiv:2102.13013 (2021).

[77] M. Kendall, L. Milsom, L. Abeler-Dörner, C. Wymant, L. Ferretti, M. Briers, C. Holmes, D. Bonsall, J. Abeler, and C. Fraser, Epidemiological changes on the isle of wight after the launch of the nhs test and trace programme: A preliminary analysis, The Lancet Digit. Health **2**, e658 (2020).

[78] G. Burgio, B. Steinegger, G. Rapisardi, and A. Arenas, Homophily in the adoption of digital proximity tracing apps shapes the evolution of epidemics, Phys. Rev. Res. **3**, 033128 (2021).

[79] M. E. J. Newman, Spread of epidemic disease on networks, Phys. Rev. E **66**, 016128 (2002).

[80] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. **87**, 925 (2015).

[81] R. Albert, H. Jeong, and A.-L. Barabási, Error and attack tolerance of complex networks, Nature (Lond.) **406**, 378 (2000).

[82] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Resilience of the Internet to Random Breakdowns, Phys. Rev. Lett. **85**, 4626 (2000).

[83] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, UK, 2008).

[84] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Critical phenomena in complex networks, Rev. Mod. Phys. **80**, 1275 (2008).

[85] E. Kenah and J. C. Miller, Epidemic percolation networks, epidemic outcomes, and interventions, Interdisc. Perspect. Infect. Dis. **2011**, 1 (2011).

[86] L. Basnarkov, M. Mirchev, and L. Kocarev, Random walk with memory on complex networks, Phys. Rev. E **102**, 042315 (2020).

[87] P. Shu, Effects of memory on information spreading in complex networks, in *Proceedings of the IEEE 17th International Conference on Computational Science and Engineering* (IEEE, Los Alamitos, CA, 2014), pp. 554–556.

[88] J. C. Miller, Epidemic size and probability in populations with heterogeneous infectivity and susceptibility, Phys. Rev. E **76**, 010101(R) (2007).

[89] E. Kenah and J. M. Robins, Second look at the spread of epidemics on networks, Phys. Rev. E **76**, 036113 (2007).

[90] P. Grassberger, On the critical behavior of the general epidemic process and dynamical percolation, Math. Biosci. **63**, 157 (1983).

[91] B. Bollobás and B. Béla, *Random Graphs*, 73 (Cambridge University Press, Cambridge, UK, 2001).

[92] F. Chung and L. Lu, The average distances in random graphs with given expected degrees, Proc. Natl. Acad. Sci. USA **99**, 15879 (2002).

[93] F. Chung and L. Lu, Connected components in random graphs with given expected degree sequences, Ann. Combinator. **6**, 125 (2002).

[94] J. C. Miller and A. Hagberg, Efficient generation of networks with given expected degrees, in *Proceedings of the International Workshop on Algorithms and Models for the Web-Graph* (Springer, Berlin, 2011), pp. 115–126.

[95] J. P. Gleeson, Cascades on correlated and modular random networks, Phys. Rev. E **77**, 046117 (2008).

[96] S. Melnik, M. A. Porter, P. J. Mucha, and J. P. Gleeson, Dynamics on modular networks with heterogeneous correlations, Chaos **24**, 023106 (2014).

[97] A. Vazquez, Multitype branching and graph product theory of infectious disease outbreaks, Phys. Rev. E **103**, L030301 (2021).