Message-passing approach to epidemic tracing and mitigation with apps

Ginestra Bianconi ⬤

School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom
and The Alan Turing Institute, 96 Euston Rd., London NW1 2DB, United Kingdom

Hanlin Sun ⬤

School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom

Giacomo Rapisardi

Barcelona Supercomputing Center (BSC) and Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili,
43007 Tarragona, Spain

Alex Arenas ⬤

Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, 43007 Tarragona, Spain

With the hit of new pandemic threats, scientific frameworks are needed to understand the unfolding of the epidemic. The use of mobile apps that are able to trace contacts is of utmost importance in order to control new infected cases and contain further propagation. Here we present a theoretical approach using both percolation and message-passing techniques, to the role of contact tracing, in mitigating an epidemic wave. We show how the increase of the app adoption level raises the value of the epidemic threshold, which is eventually maximized when high-degree nodes are preferentially targeted. Analytical results are compared with extensive Monte Carlo simulations showing good agreement for both homogeneous and heterogeneous networks. These results are important to quantify the level of adoption needed for contact-tracing apps to be effective in mitigating an epidemic.

Percolation theory [1–5] constitutes a subject of major relevance in the field of complex networks. It provides a simple mathematical framework which naturally applies to both networks' structural properties (such as resilience under random damage) [6–8], and critical diffusion (such as epidemic spreading in heterogeneous structures) [9,10]. As a matter of fact, even though there exist several epidemiological models with different flavors of complexity, the arguably most popular one, i.e., the Susceptible-Infected-Recovered (SIR) model, was found [9,10] to be mappable to a static link-percolation problem, which allowed one to find analytical expressions for the epidemic threshold depending on the underlying network topology. These results, even if they might be only an approximation of observed features in real epidemics, still constitute a fundamental theoretical cornerstone in the field of epidemic processes. Recently there has been an increasing interest in studying the effectiveness of track and tracing policies as a measure to contain epidemic spreading [11–14]: for instance, in [14] the authors show how an effective contact tracing

strategy in scale-free networks can reduce the probability of superspreading events, while in [11] it is claimed that a widely used contact-tracing app, combined with additional measures such as social distancing might be sufficient to stop an epidemic diffusion.

There are several mathematical arguments proposed in the contemporary literature to justify the above-mentioned effects; for instance in [14] a simple generating function argument is proposed in order to compute the probability that contact tracing stops the epidemic propagation; however, a solid percolation approach able to capture analytically the impact of a diffused tracing app on the nonlinear aspect of epidemic spreading has not been proposed so far. In this work, we take a step forward in filling this gap by proposing a stylized model for epidemic spreading with contact-tracing and testing policies based on link percolation.

In particular, we first consider each individual i, of a given contact network, to be assigned a binary variable T_i representing whether or not the individual has the tracing app. Then, we propose a modified version of the popular message-passing (MP) equations [15–22] which takes into account the following rationale. Every infected individual with probability p, called the *transmissibility* of the epidemic, transmits the disease to a susceptible neighbor. An individual who has the app, will know almost instantaneously (this is a hypothesis far from reality, but simplifies the analysis) if she/he has been
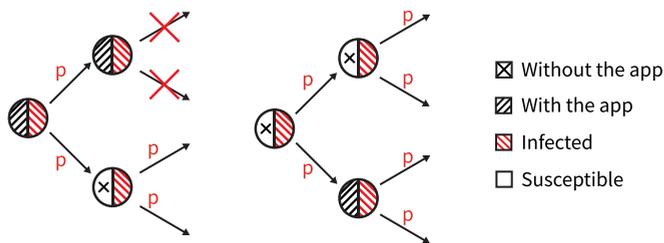
FIG. 1. Sketch of the infection pathways that lead to epidemic spreading in a population in which there are individuals that have adopted the app and individuals that have not adopted the app.

in contact with an infected individual also having the app, and she/he immediately self-isolates stopping propagation. However, if infected from an individual still not having the app, she/he will not know until symptoms appear. This can be formulated as follows: individuals with the app ($T_i = 1$) can infect only if previously infected by individuals without the app ($T_i = 0$), while individuals without the app can infect regardless of the T_i value of their infector. By doing so we are able to derive a modified nonbacktracking matrix [16,23–26] whose largest eigenvalue determines the epidemic threshold p_c. Furthermore, for the case of uncorrelated networks, we are also able to derive an analytical expression for p_c as a function of the average distribution of the tracing app, namely, $T(k)$. Our results show that in general the more the app is diffused among the population the higher is the value of p_c, meaning that the endemic state is less likely to be achieved. Moreover, we show that given a fixed app coverage on a random network ensemble, the optimal $T(k)$ which maximizes p_c corresponds to a hub-targeting strategy. By applying the message-passing algorithm to real networks, we also show that this strategy gives excellent results compared with other state-of-the-art ranking algorithms for the centrality of nodes in epidemic spreading.

Basic model of spreading with app. Let us assume a contact network $G(V, E)$ formed by $|V| = N$ individuals $i = 1, 2, \ldots, N$; each individual $i \in V$ is assigned a variable T_i indicating whether the individual has the app $T_i = 1$ or not $T_i = 0$. Assuming the contact-tracing app has an immediate effect on quarantining suspicious cases, a person with the app can infect only if she/he is infected by a person without the app, while a person without the app can infect regardless if she/he has gotten the infection from a person with the app or without the app (see Fig. 1). Now, we propose a stochastic infection model as follows: for every link (i, j) we draw a random variable $x_{ij} \in \{0, 1\}$ indicating whether the eventual contact between one infected and one susceptible node, found at the two ends of the link, leads to the infection. We parametrize this dynamic by taking $\langle x_{ij} \rangle = p$, where p indicates the transmissibility of the epidemic.

We can simulate the stationary state of this spreading process on networks of arbitrary topology, i.e., including spatial networks with high clustering coefficient, by implementing the following Monte Carlo algorithm which takes advantage of the mapping between epidemic spreading and percolation. We name T-T the links connecting two individuals adopting the app. These links do not contribute to the propagation of the

infection to nodes other than the two connected nodes. In other words, the causal chains of infection stop when they involve a T-T link. Therefore we first consider the giant component of the link percolation process in which all the T-T links are removed and all the other links are retained only if $x_{ij} = 1$. To calculate the total fraction of infected individuals in addition to the nodes in this giant component we include also the nodes with the app infected by nodes with the app [see Supplemental Material (SM) [27] for details].

Message-passing approach. To analytically predict the propagation of the epidemic on a network we use the powerful MP approach [16–18,28,29]. Although this approach is proven to give exact results only on locally treelike networks, it is also well known to be very robust in the case of networks with loops, when the underlying MP algorithm converges [30]. In this work we adopt the MP approach and we use it to predict the phase diagram of the spreading process on network ensembles as a function of the level of adoption of the app in the population.

The considered spreading model is stochastic and has different sources of randomness that can be taken into account by different MP algorithms in which we average different levels of information [17]. The simplest message MP can be derived assuming to know everything about the spreading dynamics. This would entail first to know the contact network, secondly to know which individuals have the app, i.e., the configuration $\{T_i\}_{i \in V}$, and finally to know which links have led to an actual infection, i.e., $\{x_{ij}\}_{(i,j) \in E}$ (see SM [27] for details). One can then relax the hypothesis of perfect knowledge about the epidemic process and we can consider the message passing processes in which we average over the distribution of $\{x_{ij}\}_{(i,j) \in E}$. In this situation the outcome of the epidemic spreading is dictated by the following MP equations. A node i spread the virus to node j only with probability $\sigma_{i \to j} \in [0, 1]$ where this message is found by the MP equation

$$\sigma_{i \to j} = p T_i \left[1 - \prod_{\ell \in N(i) \setminus j} [1 - (1 - T_\ell)\sigma_{\ell \to i}] \right]$$
$$+ p(1 - T_i) \left[1 - \prod_{\ell \in N(i) \setminus j} (1 - \sigma_{\ell \to i}) \right], \quad (1)$$

where $N(i)$ indicates the neighbors of node i. These equations directly implement the model as described in Fig. 1. Moreover, a node i is infected with probability $\sigma_i \in [0, 1]$ with

$$\sigma_i = \left[1 - \prod_{\ell \in N(i)} (1 - \sigma_{\ell \to i}) \right]. \quad (2)$$

Therefore the expected fraction S of infected individuals is given by

$$S = \frac{1}{N} \sum_{i=1}^{N} \sigma_i. \quad (3)$$

This process has an epidemic threshold achieved when the maximum eigenvalue $\Lambda(\mathcal{B})$ of the modified nonbacktracking matrix $\mathcal{B}$ is equal to 1, i.e.,

$$\Lambda(\mathcal{B}) = 1. \quad (4)$$

The modified nonbacktracking matrix $\mathcal{B}$ for this algorithm is defined in terms of the nonbacktracking matrix $\mathcal{A}$ of the network as

$$\mathcal{B}_{\ell i \to ij} = p(1 - T_i T_\ell) \mathcal{A}_{\ell i \to ij}. \tag{5}$$

Here $\mathcal{A}$ [16] has elements

$$\mathcal{A}_{\ell i \to ij} = a_{\ell i} a_{ij} (1 - \delta_{\ell j}), \tag{6}$$

where **a** is the adjacency matrix of the network and δ_{rs} is the Kronecker delta. Equations (4) and (5) clearly show that the epidemic threshold is dictated essentially by the nonbacktracking matrix of the network where we have removed all the T-T links.

We can also average over the probability distribution of $\{T_i\}_{i \in V}$. Specifically we can assume that $\overline{T_i}$ (the $\overline{\cdots}$ indicates the average over the probability distribution of $\{T_i\}_{i \in V}$) is only a function of the node degree, i.e., $\overline{T_i} = T(k_i)$. This is a minimal assumption that allows one to derive analytical calculations; however, we note that the adoption of the app might depend on an additional social contagion process of awareness behavior in a scenario close to the one proposed in Ref. [31]. In order to mimic these alternative scenarios in the SM [27] we have considered the cases in which the adoption of the app depends on either the eigenvector centrality or the nonbacktracking centrality of the nodes.

For formulating the MP algorithm in the case in which we assume to know only the function $T(k)$, the transmissibility p, and the actual contact network, we consider for every ordered pair of linked nodes (i, j) the two messages indicating the probability that node i infects node j given that node i has adopted ($\hat{\sigma}^T_{i \to j}$) or not adopted ($\hat{\sigma}^N_{i \to j}$) the app. These two messages are given by

$$\hat{\sigma}^T_{i \to j} = \overline{T_i \sigma_{i \to j}},$$
$$\hat{\sigma}^N_{i \to j} = \overline{(1 - T_i) \sigma_{i \to j}}. \tag{7}$$

The MP equations for these messages can be obtained by averaging the MP equations (1) over all the configuration $\{T_i\}_{i \in V}$ and read

$$\hat{\sigma}^N_{i \to j} = p \left[1 - T(k_i) \right] \left[1 - \prod_{\ell \in N(i) \setminus j} \left(1 - \hat{\sigma}^N_{\ell \to i} - \hat{\sigma}^T_{\ell \to i} \right) \right],$$

$$\hat{\sigma}^T_{i \to j} = p T(k_i) \left[1 - \prod_{\ell \in N(i) \setminus j} \left(1 - \hat{\sigma}^N_{\ell \to i} \right) \right]. \tag{8}$$

The probability that node i is infected σ_i is given by

$$\sigma_i = \left[1 - \prod_{\ell \in N(i)} \left(1 - \hat{\sigma}^N_{\ell \to i} - \hat{\sigma}^T_{\ell \to i} \right) \right], \tag{9}$$

while the expected fraction S of infected nodes is given by Eq. (3). In this case the relevant matrix $\mathcal{B}$ determining the epidemic threshold given by Eq. (4) is (see SM [27] for details)

$$\mathcal{B}_{\ell' \ell \to ij} = p[1 - T(k_i)] \delta_{\ell i} \mathcal{A}_{\ell' i \to ij}$$
$$+ p^2 [1 - T(k_i)] T(k_\ell) \mathcal{A}_{\ell' \ell \to \ell i} \mathcal{A}_{\ell i \to ij}. \tag{10}$$

Finally we consider the case in which we do not have perfect knowledge about the network itself and can perform the average over an uncorrelated network ensemble. In this case

we have two equations: one for S'_N and one for S'_T, indicating the probability that by following a link we reach an infected individual without the app or with the app, respectively. These equations (see SM [27] for details of the derivation) read

$$S'_N = p \sum_k \frac{k P(k)}{\langle k \rangle} [1 - T(k)][1 - (1 - S'_N - S'_T)^{k-1}],$$

$$S'_T = p \sum_k \frac{k P(k)}{\langle k \rangle} [T(k)][1 - (1 - S'_N)^{k-1}]. \tag{11}$$

Here $T(k)$ indicates the probability that a node of degree k gets the app. The probability that a random node gets the infection is given by

$$S = \sum_k P(k)[1 - (1 - S'_T - S'_N)^k]. \tag{12}$$

The transition is achieved for

$$p_c = \min \left(1, \frac{1}{2\kappa_T} \left[-1 + \sqrt{1 + 4 \frac{\kappa_T}{\kappa_N}} \right] \right), \tag{13}$$

where

$$\kappa_N = \frac{\langle k(k-1)[1 - T(k)] \rangle}{\langle k \rangle},$$
$$\kappa_T = \frac{\langle k(k-1) T(k) \rangle}{\langle k \rangle}, \tag{14}$$

Optimization. The formula for p_c, provided by Eq. (13), is an increasing function of κ_T so in order to maximize p_c we need to maximize κ_T. Under the $\mathbb{L}_1$ norm

$$\sum_k P(k) T(k) = \mathcal{T}. \tag{15}$$

This optimization problem gives the discrete Heaviside step function

$$\tilde{T}(k) = \theta(k - k_c, \alpha) \tag{16}$$

taking the value $0 \leqslant \alpha = \mathcal{T} - \sum_{k > k_c} P(k) < 1$ at $k = k_c$. Therefore the optimal solution is to have all nodes of degree $k > k_c$ with 100% app adoption and the node with exactly $k = k_c$ with the maximal adoption allowed by the constraint in Eq. (15). For this choice of $T(k)$ we have checked the validity of the proposed message passing theory by comparing the results obtained by a direct implementation of the Monte Carlo algorithm predicting the fraction of nodes affected by the epidemics with the results of the MP algorithm defined in Eqs. (8), (9) finding an excellent agreement between the two, for both real and synthetic networks (see Fig. 2).

Improvement on p_c. Equation (16) tells us that in an uncorrelated random network, given a fixed app coverage $\mathcal{T}$, the best strategy in order to maximally delay the percolation transition is given by targeting the hubs. In order to verify the optimality of Eq. (16) when compared to different strategies, we considered the more general form of $T(k)$ given by

$$T(k) = \rho + (1 - \rho) \theta(k - k_c, \alpha), \tag{17}$$

where $\theta(k - k_c)$ is the discrete Heaviside step function taking the value α at $k = k_c$, and $\rho \in [0, 1]$ denotes a uniform fraction of individuals adopting the app. Thanks to Eq. (17) we are able to interpolate between a purely random strategy
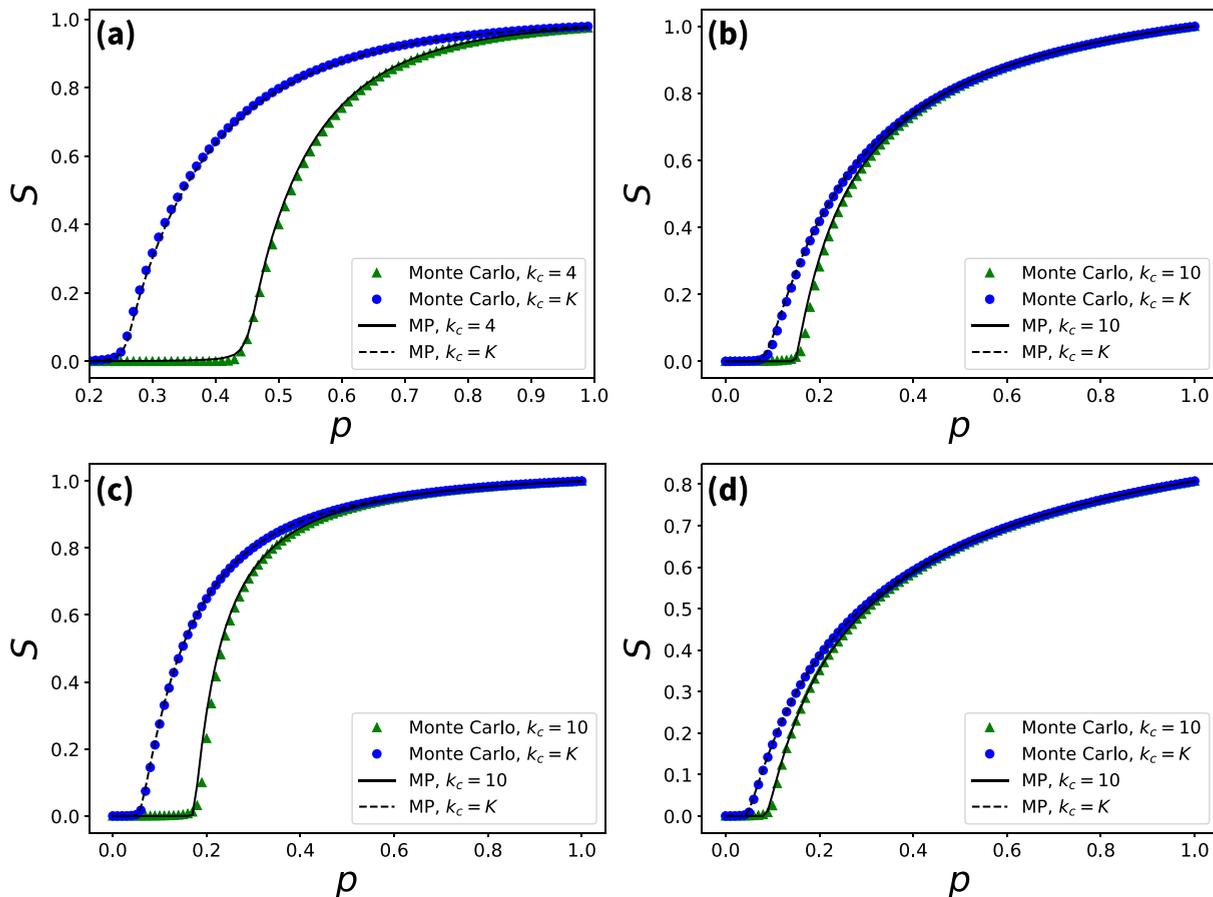
FIG. 2. The fraction of infected nodes S is plotted versus p for several networks. The results obtained by averaging the Monte Carlo simulations of the configurations $\{T_i\}_{i \in V}$ and $\{x_{ij}\}_{(i,j) \in E}$ are compared with the results of the MP algorithm defined by Eqs. (8) and (9), where $T(k)$ is given by Eq. (16) with $\alpha = 0$ and k_c as indicated in the legend of each panel. The value K in all panels corresponds to the largest degree of the network and therefore corresponds to the case of no app coverage. (a) Poisson network with $N = 5 \times 10^4$ nodes and average degree $\lambda = 4$. (b)–(d) Friendship networks from the music streaming site Deezer in the countries of Romania ($N = 41\,773$), Hungary ($N = 47\,538$), and Croatia ($N = 54\,573$), respectively [32].

obtained by taking the limit $k_c \to \infty$ and the optimal strategy given in the limit $\rho \to 0$. It is straightforward to check that under the constraint defined in Eq. (15) we have, respectively, $\lim_{k_c \to \infty} T(k) = \mathcal{T}$ and $\lim_{\rho \to 0} T(k) = \tilde{T}(k)$.

We have used Eq. (13) to investigate the phase diagram (characterized by the epidemic threshold p_c) of a Poisson network as a function of ρ and k_c (see Fig. 3). We observe that a diffused adoption of the app can significantly increase p_c, which happens when ρ increases or when k_c decreases.

To show, in a particular example, the increase of p_c due to the adoption of the app, we consider the real dataset Livemocha social network [33]. As we can see from Fig. 4, the random adoption strategy, achieved when $k_c = k_{\max}$, yields a very small increase in the value of p_c compared to the optimal distribution, corresponding to $\rho = 0$. Therefore in a scenario of limited resources, represented by the constraint defined in Eq. (15), the optimal strategy corresponds to distributing the app from higher-degree nodes to lower-degree ones until the resources are exhausted. The resulting increase in p_c computed according to Eq. (13) is quite dramatic and nontrivial; for instance from Fig. 4 we read that if the app is optimally distributed among ~40% of the population the
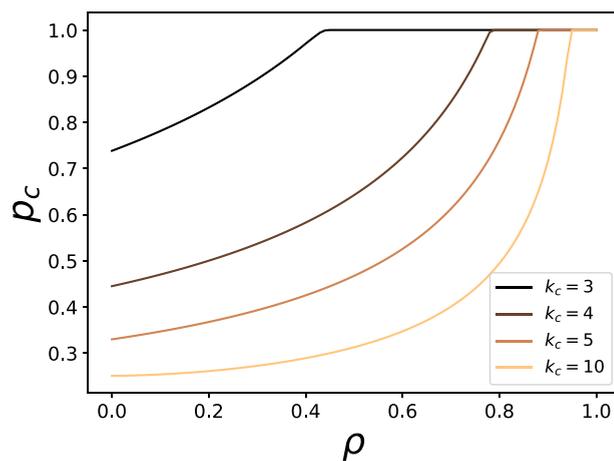

FIG. 3. The phase diagram of the epidemic model mitigated by the adoption of the app is shown for a Poisson network of $N = 10^4$ nodes with average degree $\lambda = 4$. Here $T(k)$ is given by Eq. (17) with $\alpha = 0$. The epidemic threshold p_c is plotted as a function of ρ for different values of the cutoff k_c.
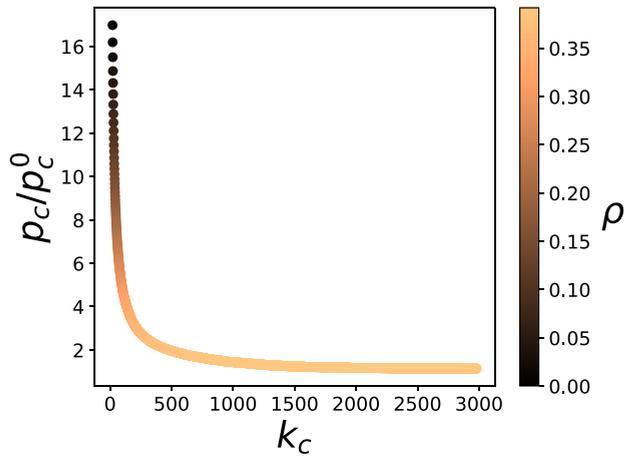
FIG. 4. Relative increase of p_c computed from Eq. (13) on the Livemocha social network ($N \sim 104 \times 10^3$ nodes, $E \sim 2 \times 10^6$ edges) [33], where $T(k)$ is given by Eq. (17) under the constraint (15), and $p_c^0 = \langle k \rangle / \langle k(k-1) \rangle$ represents the value of the percolation threshold in the absence of app coverage [which can be obtained from Eq. (13) in the limit $\kappa_T \to 0$]. Here $p_c^0 = 0.003\,06$, while the app coverage is fixed at $\mathcal{T} = 0.391\,75$, corresponding to an optimal $\bar{T}(k)$ with $k_c = 20$ and $\alpha = 1$. The plot shows that for this particular value of $\mathcal{T}$, corresponding to $\sim 40\%$ of the nodes having the app, the optimal distribution is reached at $\rho = 0$ and corresponds to an ~ 17-fold increase of p_c, whereas in the case of a purely random strategy, obtained at $\rho = \mathcal{T}$, the increase of p_c is ~ 1.2-fold.

increase of p_c is ~ 17-fold, while if the same percentage is covered randomly the increase is ~ 1.2-fold. This optimization principle is obtained under the assumption that the adoption of the app is dictated by the degree of the nodes. However, in a real scenario this hypothesis might appear too restrictive. Devising an *ad hoc* optimization algorithm similar to the ones proposed in [28,29,34] is beyond the scope of this Letter. However, in order to check how the obtained optimal strategy compares with other possible mechanisms driving the adoption of the app in the SM we show that targeting the hubs remains a very good strategy also if compared to targeting the high eigenvector centrality nodes or the high nonbacktracking centrality nodes [25,26] in a number of real datasets.

Conclusions. In this work we provide a message-passing theory able to predict the epidemic threshold of disease spreading among a population which has the option of adopting a tracing app. The simplicity of our model allows us to derive a simple analytical estimate for the epidemic threshold

and leaves plenty of room for taking into account more complex and realistic factors. For instance, we assumed that the tracing app is perfect; however, we can relax this assumption in order to allow also for imperfect tracing and isolation. Another interesting followup for the model could be the introduction of a time dimension similar to the one proposed in [25], in order to assess how the modified nonbacktracking matrix presented in Eq. (5) affects not only the percolation threshold itself, but also the speed of the epidemic.

The proposed stylized mathematical framework can overall be useful to assess the expected impact of contact-tracing apps in the course of an epidemic if adopted correctly. The compartmental epidemic model used is the classical SIR, and does not pretend to be a model fitted for the current pandemic of COVID-19; however, the physical intuition we grasp from the presented analysis may prove fundamental to prescribe the best targeting strategy for app adoption, as well as it captures the highly nonlinear effect on the reduction of the incidence provided by a certain fraction of adoption. Our preliminary results show both numerically and theoretically that the adoption of the app by a large fraction of the population increases the value of the epidemic threshold. In the case of uncorrelated networks we are able to derive a closed analytic expression for p_c which depends on both the network degree distribution $P(k)$ and the average app distribution $T(k)$. Thanks to this expression we finally prove in a constrained-resources scenario that the value of p_c is maximized when high-degree nodes are preferentially targeted. Our results show that an optimal targeting gives rise to a dramatic increase in the value of p_c when compared to a strategy in which the same amount of resources is uniformly distributed. The more randomly the app is diffused among the population the less is the increase in the percolation threshold, or equivalently, the less the app has the power of mitigating the epidemics. Overall our results show that even if the adoption of a tracing app has the effect of preventing an epidemic wave, the same level of adoption can be optimally distributed by taking into account the heterogeneity of the population contact network in order to obtain a mitigation effect which is significantly higher.

[1] A.-L. Barabási, *Network Science* (Cambridge University Press, Cambridge, UK, 2016).

[2] M. Newman, *Networks: An Introduction* (Oxford University Press, New York, 2018).

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

[4] R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, Cambridge, UK, 2010).

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

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

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

[8] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Structure of Growing Networks with

Preferential Linking, Phys. Rev. Lett. **85**, 4633 (2000).

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

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

[11] 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).

[12] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. Pastore y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M. E. Halloran, I. M. Longini, and A. Vespignani, The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak, Science **368**, 395 (2020).

[13] 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).

[14] S. Kojaku, L. Hébert-Dufresne, and Y.-Y. Ahn, The effectiveness of contact tracing in heterogeneous networks, arXiv:2005.02362 (2020).

[15] B. Karrer and M. E. J. Newman, Message passing approach for general epidemic models, Phys. Rev. E **82**, 016101 (2010).

[16] B. Karrer, M. E. J. Newman, and L. Zdeborová, Percolation on Sparse Networks, Phys. Rev. Lett. **113**, 208702 (2014).

[17] G. Bianconi, *Multilayer Networks: Structure and Function* (Oxford University Press, New York, 2018).

[18] F. Radicchi and G. Bianconi, Redundant Interdependencies Boost the Robustness of Multiplex Networks, Phys. Rev. X **7**, 011013 (2017).

[19] G. T. Cantwell and M. E. J. Newman, Message passing on networks with loops, Proc. Natl. Acad. Sci. USA **116**, 23398 (2019).

[20] F. Altarelli, A. Braunstein, L. Dall'Asta, J. R. Wakeling, and R. Zecchina, Containing Epidemic Outbreaks by Message-Passing Techniques, Phys. Rev. X **4**, 021024 (2014).

[21] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, Vital nodes identification in complex networks, Phys. Rep. **650**, 1 (2016).

[22] S. Mugisha and H.-J. Zhou, Identifying optimal targets of network attack by belief propagation, Phys. Rev. E **94**, 012305 (2016).

[23] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, Spectral redemption in clustering sparse networks, Proc. Natl. Acad. Sci. USA **110**, 20935 (2013).

[24] T. Martin, X. Zhang, and M. E. J. Newman, Localization and centrality in networks, Phys. Rev. E **90**, 052808 (2014).

[25] S. Moore and T. Rogers, Predicting the Speed of Epidemics Spreading in Networks, Phys. Rev. Lett. **124**, 068301 (2020).

[26] T. Rogers, Assessing node risk and vulnerability in epidemics on networks, Europhys. Lett. **109**, 28005 (2015).

[27] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.3.L012014 for further details on the analytical derivations, on the used algorithms and on the numerical results discussed in the main body of the paper.

[28] A. Y. Lokhov, M. Mézard, H. Ohta, and L. Zdeborová, Inferring the origin of an epidemic with a dynamic message-passing algorithm, Phys. Rev. E **90**, 012801 (2014).

[29] F. Altarelli, A. Braunstein, L. Dall'Asta, A. Lage-Castellanos, and R. Zecchina, Bayesian Inference of Epidemics on Networks Via Belief Propagation, Phys. Rev. Lett. **112**, 118701 (2014).

[30] S. Melnik, A. Hackett, M. A. Porter, P. J. Mucha, and J. P. Gleeson, The unreasonable effectiveness of tree-based theory for networks with clustering, Phys. Rev. E **83**, 036112 (2011).

[31] C. Granell, S. Gómez, and A. Arenas, Competing spreading processes on multiplex networks: Awareness and epidemics, Phys. Rev. E **90**, 012808 (2014).

[32] B. Rozemberczki, R. Davies, R. Sarkar, and C. Sutton, GEM-SEC: Graph embedding with self clustering, in *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (ACM, New York, USA, 2019), pp. 65–72.

[33] R. Zafarani and H. Liu, Social Computing Data Repository at ASU, Arizona State University, School of Computing, Informatics and Decision Systems Engineering, 2009, http://socialcomputing.asu.edu.

[34] F. Morone and H. A. Makse, Influence maximization in complex networks through optimal percolation, Nature (London) **524**, 65 (2015).