TEST ALLOCATION BASED ON RISK OF INFECTION FROM FIRST AND SECOND ORDER CONTACT TRACING

Gabriela Bayolo Soler [1] & Miraine Dávila Felipe [2] & Ghislaine Gayraud [3]

[1] *Université de technologie de Compiègne - France - gabriela.bayolo-soler@utc.fr*
[2] *Université de technologie de Compiègne - France - miraine.davila-felipe@utc.fr*
[3] *Université de technologie de Compiègne - France - ghislaine.gayraud@utc.fr*

Résumé. Face à une quantité limitée des ressources disponibles, les stratégies visant à atténuer la propagation d'une épidémie, telles que les tests aléatoires et la recherche des contacts, deviennent inefficaces. Nous proposons ici d'allouer les ressources de manière plus efficace, en calculant le risque individuel d'infection au cours du temps, basé sur l'observation partielle de la propagation de l'épidémie sur un réseau de contacts ; ce risque est défini comme la probabilité d'être infecté par n'importe quelle chaîne de transmission possible jusqu'à une longueur de deux, provenant d'individus récemment détectés. Pour évaluer les performances de notre méthode et les effets de certains paramètres clés, nous effectuons des expériences comparatives simulées en utilisant des données générées par un modèle basé sur des agents.

Mots-clés. réseaux de contacts, recherche des contacts, atténuation des épidémies, probabilité d'infection, estimation des risques, attribution des tests

Abstract. Under limited available resources, strategies for mitigating the propagation of an epidemic such as random testing and contact tracing become inefficient. Here, we propose to accurately allocate the resources by computing over time an individual risk of infection based on the partial observation of the epidemic spreading on a contact network; this risk is defined as the probability of getting infected from any possible transmission chain up to length two, originating from recently detected individuals. To evaluate the performance of our method and the effects of some key parameters, we carry out comparative simulated experiments using data generated by an agent-based model.

Keywords. contact networks, contact tracing, epidemic mitigation, probability of infection, risk estimation, test allocation

Introduction

In this work, we propose a method to compute the risk of infection of individuals in the population over time, based on the partial observation of the epidemic spreading through the population contact network. The risk of each individual is defined as her/his (marginal) probability of infection conditionally on the observed variables in the recent past, and the higher-risk individuals can get notified to be tested, quarantined, or applied any other preventive measures. Thus, the quantification of the infection risk is proposed here as a tool to allocate the available resources more rationally than just randomly. Similar intervention approaches have been shown to have a positive impact on the mitigation of epidemics, see for instance Baker et al. (2021); Herbrich et al. (2022); Batlle et al. (2022); Romijnders et al. (2023); Guttal et al. (2020); Bestvina and Thornton (2021); Murphy et al. (2021); Gupta et al. (2023); Bengio et al. (2021); Sattler et al. (2020); Alsdurf et al. (2020).

Among this recent research with the same aim as ours, in Bestvina and Thornton (2021) and Batlle et al. (2022), the risk is computed using Monte Carlo methods. In Bestvina and Thornton (2021), the authors estimate the individual infection probability up to 3° contact tracing, arguing that it improves the detection of asymptomatic patients in diseases with a high percentage of them. Here, instead, we derive an explicit formula for these probabilities taking into account up to 2° contacts, and hence avoiding the large computing power and the centralised information required by Monte Carlo methods. Other works such as Baker et al. (2021) and Guttal et al. (2020) avoid the use of Monte Carlo methods by using the mean-field approximation to evaluate the individual risk of infection. However, the way the risk is propagated is "bidirectional", meaning that it is not only "forward" in the direction of the transmission given the observations. Indeed, they suppose that individuals interchange their risk information at each time step if there is an edge between them, regardless of the previous path followed by the transmitted risk. Despite the similarities in the use of the MF hypothesis, it should be noticed that in Guttal et al. (2020) the network and propagation model are simpler than in Baker et al. (2021). The latter deals with more realistic models, including the OpenABM-Covid19 model used in our simulations. Moreover, in Baker et al. (2021) a second method is developed, that estimates the individual infection risk as the posterior distribution conditional on the test observations through the Belief Propagation inference algorithms. Similar computations are achieved in Herbrich et al. (2022) and Romijnders et al. (2023) using Gibbs Sampling and Factorized Neighbors respectively. Another related methodology is presented in a series of works Bengio et al. (2021); Gupta et al. (2023); Alsdurf et al. (2020); Biazzo et al. (2022); Shah et al. (2020); Čutura et al. (2021); Tomy et al. (2022); Tan et al. (2023), where the authors achieve the risk computation using deep learning algorithms based on neural networks.

We propose and simulate the following mitigation strategy: every day the probability of being infected is computed for the individuals at risk, and a fixed number (η, number of daily available tests) of the highest-ranked individuals are tested; the newly detected individuals are put in quarantine the day after, and the process is repeated each day during an intervention period.

We briefly describe our approach in Section 1 and we present some results in Section 2.

1 Methods

To be more precise about our approach, we suppose that the infectious statuses of individuals are (partially) observed through testing, as well as the underlying contact network. Actually, we consider *at risk* not only the first-degree contacts of detected individuals (1° contacts) but also their subsequent contacts (2° contacts). In the sequel, to distinguish the different groups of individuals under consideration, we call *index cases* the detected individuals who are infectious, 1° *contacts* the individuals who are not detected but interacted with index cases while the latter were infectious, and, 2° *contacts* the individuals who are not detected but interacted with 1° contacts after the latter were in contact with index cases. In addition, 1° *interaction* and 2° *interaction* refer to the risky encounter between index cases and 1° contacts, and between 1° and 2° contacts respectively. Then, we compute the probability for each individual at risk of having been infected by one of the index cases in the previous days (fixed time window), through chains of transmission of length one or two.

Infectious disease spreading on the network

We consider a population consisting of N ($N \in \mathbb{N}$) individuals that stays constant over time. At any discrete time t ($t \in \mathbb{N}$), the social structure (interactions between individuals at t) is represented by an undirected graph $\mathcal{G}_t = (\mathcal{V}, \mathcal{E}_t)$, corresponding to the set of vertices $\mathcal{V}$ (the individuals) and the set of the edges $\mathcal{E}_t$ (the interactions at time t between the corresponding individuals). Here, we consider an individual-based SIR dynamic spreading on the underlying social network. The possible individual statuses are only Susceptible (S), Infected (I) and Removed (R), and the only possible status evolution over time are $S \to I$ and $I \to R$, where R considered as an absorbing state. We denote by $X_t^i \in \{S, I, R\}$ the random variable corresponding to the status of individual i at time t.

Observations

At a given time $t \geq 0$, the set of observations $\mathcal{O}_t$ is provided by the graph of interactions during the recent days and the set of individuals with a positive and a negative result.

Risk of infection via transmission chains

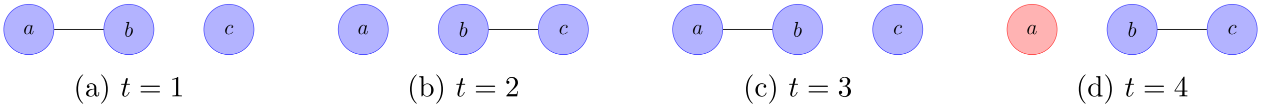
For any time t and any individual j at risk, our aim is to estimate the probability of infection of j given the set of observations $\mathcal{O}_t$, that is

$$\mathbb{P}\left(X_t^j = I \;\middle|\; \mathcal{O}_t\right) \tag{1}$$

In order to compute this probability we propose two approaches based on two different degrees of interactions. To differentiate both methods, we call them in the sequel 1°contact tracing (1°CT) and 2°contact tracing (2°CT). In the first approach, the risk of infection is based on 1° interactions, while the second proposes a more accurate risk of infection, defined from both 1° and 2° interactions.

We illustrate, using the toy example in Figure 1, the fundamental differences between 1°CT and 2°CT method. The Figure shows the interactions between 3 individuals during 4 days. At day 4, individual a is tested positive and then their contacts in the past are traced and ranked. For 1°CT is considered only individual b at risk, while 2°CT considers both individuals b and c.

We introduce a truncation parameter $\gamma \in \mathbb{N}$ corresponding to the infection time-frame of interest. More precisely, for any individual not tested positive, we are interested in the



(a) $t = 1$ (b) $t = 2$ (c) $t = 3$ (d) $t = 4$

Figure 1: Temporal interaction networks in a population of $N = 3$. Individual a is detected at day $t = 4$.

approximate probability of being infected in the interval $[t-\gamma : t]$ given the set of observations at time t. For 2°CT method, we also introduce the parameter ζ ($\zeta \geq \gamma$) as the time-frame for the infection date of the 1°contacts, meaning that it lies in the interval of time $[t-\zeta : t-1]$. As a consequence, we are interested in 1° interactions that occur in $[t-\zeta : t-1]$ and in the 2° interactions that occur after a possible transmission due to a 1° interaction in the interval of time $[t-\gamma : t]$.

2 Simulation results

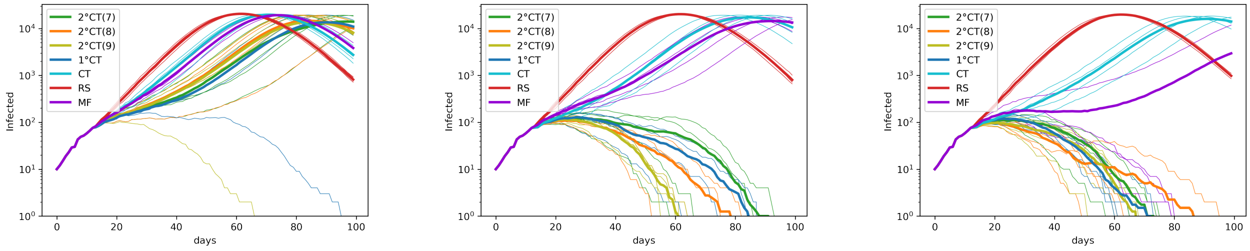
The simulation starts at time $t = 0$. At the beginning, all individuals are susceptible (S), except for a small number N_0 of infectious individuals ("patients zero"). Every day, starting from $t = 1$, a proportion p_s, respectively p_m, of individuals with newly developed severe and mild symptoms are tested, detected and quarantined. Later, at a fixed date t_0 in $[1 : T]$, the intervention starts, and it is carried out daily until the mitigation of the epidemic or the end of the study. At any $t \geq t_0$, the intervention strategy based on the 2°CT method consists of tracing 1° and 2°contacts, computing their risk of infection, and ranking them according to their risk values. Then, the first η individuals in the ranking are tested, and the newly detected ones become index cases and are quarantined. The default quarantine protocol stops the interactions in the occupation and random network, but those within the household are maintained. The tests are assumed to be perfect, and we suppose that the test results are available the same day on which the tests are performed. The intervention related to the 1°CT method is analogous to the one for the 2°CT method, except that only the 1°contacts are traced and ranked.

To test our proposed method on a proper data set, we generate the data using the OpenABM-Covid19 model introduced by Hinch et al. (2021).

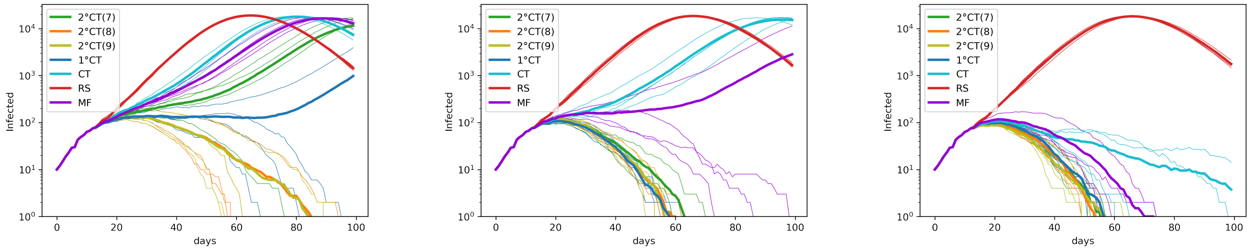
Results

In this section, we present the results obtained using the intervention based on the risk, through the simulation of different scenarios. For all the simulations, the propagation of the epidemic is identical until t_0, while it might change after t_0, when the intervention method starts, depending on the particular scenario. In the figures each thin line represents the result obtained for the realisation associated with one *seed*, while the thick lines correspond to the average of all the realisations.

To evaluate the efficiency of the proposed 1°CT and 2°CT methods to mitigate an epidemic, we compare them with three other ranking strategies: Random Selecting (RS, individuals are ranked randomly), Contact Tracing (CT, individuals are ranked according to their number of interactions with detected individuals in the time-frame $[t - \gamma : t]$) and Mean-Field (MF, individuals are ranked according to the mean-field risk approximation presented in Baker et al. (2021)).



(a) $\eta = 125, p_m = 0.5$ (b) $\eta = 250, p_m = 0.5$ (c) $\eta = 400, p_m = 0.5$



(d) $\eta = 125, p_m = 0.75$ (e) $\eta = 250, p_m = 0.75$ (f) $\eta = 400, p_m = 0.75$

Figure 2: Effect of the parameters η (the number of daily available tests, increasing from left to right) and p_m (the proportion of daily detected individuals with mild symptoms, increasing from top to bottom) on the epidemic spreading for the strategies 1°CT, 2°CT, CT, RS and MF. In all simulations we consider $T = 100$, $N = 50K$, $t_0 = 12$, $N_0 = 10$ and $p_S = 1$. We fix the parameters $\gamma = 6$ and $\zeta = 7, 8, 9$ (indicated in the legend as 2°CT(7), 2°CT(8), 2°CT(9), respectively). The values for the parameters in MF strategy are $\rho_{MF} = 5$ and $t_{MF} = 10$.

We compare the five strategies in Figure 2, in which we display the number of infectious individuals in logarithmic scale through time across a broad range of values for the parameters. In particular, we increase the number of daily available tests from the left panels to the right ones, and we increase the proportion of daily detected mild symptomatic individuals from top to bottom. As expected for all strategies, a higher value of p_m and/or η improves the mitigation of the epidemic in terms of the duration and the total number of infected individuals. The simulations show that our proposed methods (1°CT and 2°CT) improve considerably the results compared to the MF and the usual CT, which are all better than the RS strategy. The latter does not mitigate the epidemic even with a high number of daily available tests and a high value of p_m, while the MF and CT methods achieve the mitigation for a large value of η. We also study the 2°CT method for different time-frames in which the 1°contact can get infected, that is in $[t - \zeta : t]$, where we consider $\zeta = 7$ in green, $\zeta = 8$ in orange and $\zeta = 9$ in yellow. Figure 2 shows that the results are improved as ζ increases. In particular, it should be noticed that the 2°CT method with $\zeta = 8$ and $\zeta = 9$ gives better results than the 1°CT method. However, the 1°CT method requires less individual information and therefore it is better in terms of privacy restrictions. From these results, a trade-off can arise between getting better results with computationally demanding (2°CT method) and preserving individual privacy with simpler and faster computation. Indeed, it is worth mentioning that for a high enough number of daily available tests and/or a high enough proportion of mild observed, the methods 1°CT and 2°CT have similar effects on the mitigation of the epidemic; hence in this case, we recommend the use of the 1°CT method than the 2°CT method.

Conclusions

A crucial aspect of our work is that we consider a rather realistic contact network and a detailed disease spread model, see Hinch et al. (2021). Another significant aspect is that we consider 2° contact tracing, providing a more accurate estimation of the risk compared with the 1° contact tracing. Although our approach can be extended to 3° contacts and beyond, the calculations would get much heavier, and we argue that the gain in the effectiveness of the mitigation would not be significant, due to the uncertainty on the statuses of the intermediate individuals in the chains of transmission. One more core feature of our method is that to compute the risk, neither the whole contact network nor a centralised setup (contacts and individual information) is required. These characteristics are essential for the practical implementation using digital contact tracing applications, where the interchange of information between contacts over the whole network could be difficult due to a huge amount of personal data impacted by privacy restrictions. These difficulties can be intensified in the centralised case, see Romijnders et al. (2023). Furthermore, compared to the previously inference algorithms in the literature used to calculate the individual risk of being infected (Belief Propagation, Gibbs Sampling and Factorized Neighbors), our risk calculation is simpler: while these algorithms integrate the observations at any time t by updating and re-propagating the risks step-by-step in a given time interval previous to t for every contact (up to any contact degree) of all the individuals in the population, we calculate directly the risk at t of individuals in contact (up to 2°) with someone detected by integrating the probability of any possible path of length up to 2 that might lead to the infection of these individuals. In this way, we avoid any cycling back phenomenon, and we do not need to update the risk of all individuals for every time step in the contact tracing time window, getting a very low level of messages interchange between individuals, see Romijnders et al. (2023).

Further information

The repository with the code is available in https://github.com/gbayolo26/risk_estimation. The presentation will be based on the recent submitted paper Bayolo Soler et al. (2023).

References

Alsdurf, H., Belliveau, E., Bengio, Y., Deleu, T., Gupta, P., Ippolito, D., Janda, R., Jarvie, M., Kolody, T., Krastev, S., Maharaj, T., Obryk, R., Pilat, D., Pisano, V., Prud'homme, B., Qu, M., Rahaman, N., Rish, I., Rousseau, J.-F., Sharma, A., Struck, B., Tang, J., Weiss, M. and Yu, Y. W. (2020) Covi white paper.

Baker, A., Biazzo, I., Braunstein, A., Catania, G., Dall'Asta, L., Ingrosso, A., Krzakala, F., Mazza, F., Mé zard, M., Muntoni, A. P., Refinetti, M., Mannelli, S. S. and Zdeborová, L. (2021) Epidemic mitigation by statistical inference from contact tracing data. *Proceedings of the National Academy of Sciences*, **118**, e2106548118.

Batlle, P., Bruna, J., Fernandez-Granda, C. and Preciado, V. M. (2022) Adaptive test allocation for outbreak detection and tracking in social contact networks. *SIAM Journal on Control and Optimization*, **60**, S274–S293.

Bayolo Soler, G., Dávila Felipe, M. and Gayraud, G. (2023) Test allocation based on risk of infection from first and second order contact tracing. *arXiv preprint arXiv:2311.02094*.

Bengio, Y., Gupta, P., Maharaj, T., Rahaman, N., Weiss, M., Deleu, T., Muller, E. B., Qu, M., Schmidt, V., St-Charles, P.-L., Alsdurf, H., Bilaniuk, O., Buckeridge, D., Marceau-Caron, G., Carrier, P.-L., Ghosn, J., Ortiz Gagne, S., Pal, C., Rish, I., Schölkopf, B., Sharma, A., Tang, J. and Williams, A. (2021) Predicting infectiousness for proactive contact tracing. In *9th International Conference on Learning Representations (ICLR)*. Virtual Conference.

Bestvina, I. and Thornton, W. (2021) Contact tracing infection risk estimate. Distributed simulation aproach. URL: `https://www.viratrace.org/#/`.

Biazzo, I., Braunstein, A., Dall'Asta, L. and Mazza, F. (2022) A bayesian generative neural network framework for epidemic inference problems. *Scientific Reports*, **12**, 19673.

Gupta, P., Maharaj, T., Weiss, M., Rahaman, N., Alsdurf, H., Minoyan, N., Harnois-Leblanc, S., Merckx, J., Williams, A., Schmidt, V. et al. (2023) Proactive contact tracing. *PLOS Digital Health*, **2**, e0000199.

Guttal, V., Krishna, S. and Siddharthan, R. (2020) Risk assessment via layered mobile contact tracing for epidemiological intervention. *medRxiv*. URL: `https://www.medrxiv.org/content/early/2020/05/01/2020.04.26.20080648`.

Herbrich, R., Rastogi, R. and Vollgraf, R. (2022) CRISP: A Probabilistic Model for Individual-Level COVID-19 Infection Risk Estimation Based on Contact Data.

Hinch, R., Probert, W. J., Nurtay, A., Kendall, M., Wymant, C., Hall, M., Lythgoe, K., Bulas Cruz, A., Zhao, L., Stewart, A. et al. (2021) OpenABM-Covid19—an agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. *PLoS computational biology*, **17**, e1009146.

Murphy, K., Kumar, A. and Serghiou, S. (2021) Risk score learning for COVID-19 contact tracing apps. In *Proceedings of the 6th Machine Learning for Healthcare Conference*, vol. 149 of *Proceedings of Machine Learning Research*, pp. 373–390. PMLR. URL: `https://proceedings.mlr.press/v149/murphy21a.html`.

Romijnders, R., Asano, Y. M., Louizos, C. and Welling, M. (2023) No time to waste: practical statistical contact tracing with few low-bit messages. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, vol. 206 of *Proceedings of Machine Learning Research*, 7943–7960. PMLR. URL: `https://proceedings.mlr.press/v206/romijnders23a.html`.

Sattler, F., Ma, J., Wagner, P., Neumann, D., Wenzel, M., Schäfer, R., Samek, W., Müller, K.-R. and Wiegand, T. (2020) Risk estimation of SARS-CoV-2 transmission from bluetooth low energy measurements. *NPJ Digital Medicine*, **3**, 129. URL: `https://doi.org/10.1038/s41746-020-00340-0`.

Shah, C., Dehmamy, N., Perra, N., Chinazzi, M., Barabási, A.-L., Vespignani, A. and Yu, R. (2020) Finding patient zero: Learning contagion source with graph neural networks.

Tan, C. W., Yu, P.-D., Chen, S. and Poor, H. V. (2023) Deeptrace: Learning to optimize contact tracing in epidemic networks with graph neural networks.

Tomy, A., Razzanelli, M., Di Lauro, F., Rus, D. and Della Santina, C. (2022) Estimating the state of epidemics spreading with graph neural networks. *Nonlinear Dynamics*, **109**, 249–263.

Čutura, G., Li, B., Swami, A. and Segarra, S. (2021) Deep demixing: Reconstructing the evolution of epidemics using graph neural networks. In *2021 29th European Signal Processing Conference (EUSIPCO)*, 2204–2208.