

Improving contact tracing by prioritizing influential spreaders identified by socio-demographic characteristics

Marius Kaffai

University of Stuttgart, Germany, marius.kaffai@sowi.uni-stuttgart.de

Abstract. Contact tracing is an effective but costly non-pharmaceutical intervention to control infectious diseases. Aiming to increase its efficacy, we test whether contact tracing benefits from prioritizing influential spreaders identified based on socio-demographic characteristics. By utilizing proxy measures of influential spreading, this approach circumvents the common problem of targeted interventions that influential spreaders are difficult to identify in real-world scenarios. Simulation experiments executed in an empirically-calibrated agent-based model show that tracing the contacts of influential spreaders identified by age and household size is superior to strategies that prioritize random, the youngest, or the oldest agents.

Keywords: Contact tracing · Influential spreaders · Agent-based model

1 Introduction

It is frequently suggested to apply interventions focusing specifically on influential spreaders to fight a pandemic with higher efficacy [5, 10]. Because the potential for influential spreading is, in most cases, not directly observable in real-world scenarios, it is additionally suggested to determine suitable proxy measures to make such approaches more feasible [8]. Aiming to explore the potential benefits of the proposed approach, we use an empirically-calibrated agent-based model to test whether the non-pharmaceutical intervention of contact tracing could be optimized by prioritizing people with a higher risk for influential spreading estimated based on age and household size. Although there is already a high number of (modeling) studies focusing on contact tracing [9], up to our knowledge, there is no research examining the potential improvement of contact tracing by prioritizing expected influential spreaders identified based on socio-demographic characteristics.

Examining the optimization of contact tracing is especially interesting because of its high resource intensity. Although digital types of contact tracing exist, in many countries, such as Germany, manual contact tracing, in which local public health authorities interview the infected, trace their contacts and monitor quarantine compliance, is still the predominant type of contact tracing. In addition, in contrast to interventions like vaccinations, contact tracing does

not offer any protection to the infected but only to the contacts of the infected or the contacts of the contacts of the infected. Hence, known strategies like prioritizing the oldest to reduce hospitalizations, as it is effective for vaccination campaigns [6], appear to be insufficient in the context of contact tracing.

Age and household size are used as socio-demographic predictors to identify influential spreaders because they are easy to observe in the real world, thus ensuring the approach’s feasibility. In addition, there is research showing the importance of both variables regarding the vulnerability to infectious respiratory diseases [1,4], which suggests that they might also be important to the spreading of infectious diseases. Furthermore, the existing research and data allow us to partly validate the simulation results with regard to those two predictors.

The procedure of this research project is as follows: In the first step, we run an infectious disease simulation modeling the spread of a COVID-19-like disease with no intervention present. Using the generated data, we regress the number of infections caused directly or indirectly on age and household size. In the second step, we run simulation experiments to evaluate the performance of contact tracing when prioritizing contacts of agents for which the regression model predicts the highest potential for influential spreading.

Tracing only the contacts of agents with the highest predicted potential for influential spreading is superior under most tested conditions compared to three reference strategies. In general, the results show that contact tracing might benefit from identifying influential spreaders and tracing only their contacts. In particular, due to the specific calibration of the model to the characteristics of the German population and COVID-19, the results suggest that age and household size are suitable proxy measures when implementing such an approach in Germany to control COVID-19-like infectious diseases.

2 Methods

The agent-based model simulates the spread of a COVID-19-like infectious disease in a society of people living their daily life (an earlier version of the model is described in detail in [3]). The main model procedures are agents visiting various locations (e.g., home, work, or school) based on their attributes (e.g., age, work time per day), encountering other agents at those locations, infecting encountered agents, or becoming infected. To attain a detailed representation of the micro-level, we construct the population of agents directly from the data of complete households of survey participants provided by the GSOEP [2]. Hence, both the agent attributes and the closest contact network are directly determined by empirical data. To calculate the transmission probability between two agents, we create a weighted contact network based on the time each agent visits a location. To validate and calibrate the model, we use empirical data on age-specific contact patterns [7] and on cases of COVID-19 in Germany.

To obtain a predictive regression model that can be used to identify supposed influential spreaders based on age and household size, we first run an infectious disease simulation in a scenario with no interventions. We count the number

of infections caused directly or indirectly by each agent to measure influential spreading. We regress that number on age and household size in the second step using a negative binomial regression model. Because of the expected non-linearity of the effect of age, age is categorized into six groups in orientation to the age groups frequently used by the German Robert Koch Institute.

To test how contact tracing performs when prioritizing expected influential spreaders, a simulation experiment was set up. As a baseline reference, we run a scenario where contact tracing is implemented with a random selection of agents whose contacts are traced. As additional references, we implemented the prioritization of the youngest and the oldest agents. Contact tracing is implemented in the simulation model as follows: At the beginning of each run, a specific proportion of agents, which is varied in the simulation experiments, is selected. Depending on the scenario, the proportion of selected agents includes random agents, the agents with the highest predicted potential for influential spreading, the youngest or the oldest agents. During the simulation, only contacts of those selected agents will be traced. The traced contacts, which are 80% of the agents an index case had the most contact with during the three days before turning symptomatic ill, are sent to strict quarantine for the subsequent ten days.

3 Results

Concerning the identification of influential spreaders by socio-demographic attributes, the regression analysis reveals that age and household size are relevant predictors of influential spreading (Nagelkerke $R^2 = 0.38$). For each household member, the expected number of caused infections increases by a factor of 1.48 on average, holding age constant. Age shows a non-monotone effect, with middle-aged agents causing the most infections (e.g., agents aged between 15 and 34 cause 1.89 times as many infections as agents younger than 5) and older adults causing the least infections (e.g., agents between 60 and 79 cause only 0.64 times as many infections as agents younger than 5).

Turning to the results of the simulation experiments, panel A in Figure 1 shows that all strategies can reduce the number of infections substantially. However, targeting the estimated most influential spreaders is significantly more effective than other strategies for lower and medium proportions of people whose contacts are traced. As panel B in Figure 1 shows, the improvement of this strategy relative to the random scenario peaks when the contacts of about 20-30 % of the most influential spreaders are being traced in case of infection. Because the number of selected agents whose contacts are potentially traced in case of infection does not equal the number of agents traced during the simulation, panels C and D show the number of prevented infections per one contact traced. Again, the proposed strategy of targeting only the expected most influential spreaders is superior for medium and low proportions of agents whose contacts are traced. The results for hospitalizations show a very similar pattern (not shown in this abstract).

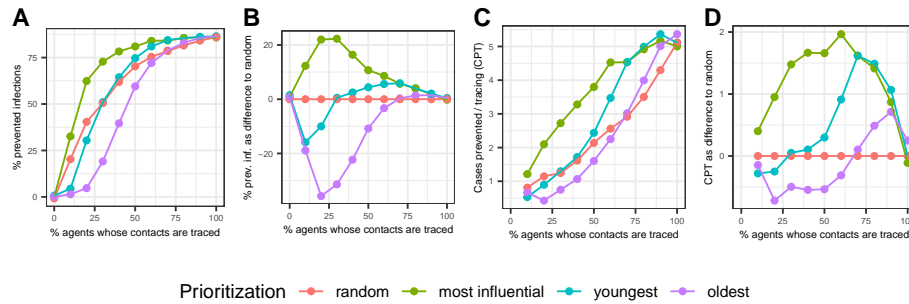


Fig. 1. **A:** Average number of infections that could be prevented relative to a scenario with no contact tracing as a function of the number of agents whose contacts were traced. **B:** Same as **A** but as difference to the scenario with no prioritization strategy (random). **C:** Average number of infections that could be prevented per one traced contact as a function of the number of agents whose contacts were traced. **D:** Same as **C** but as difference to the scenario with no prioritization strategy (random).

References

1. M. R. A. Cardoso, S. N. Cousens, L. F. de Góes Siqueira, F. M. Alves, and L. A. V. D'Angelo. Crowding: risk factor or protective factor for lower respiratory disease in young children? *BMC Public Health*, 4(1):19, June 2004.
2. J. Goebel, M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp. The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik*, 239(2):345–360, Apr. 2019.
3. M. Kaffai and R. H. Heiberger. Modeling non-pharmaceutical interventions in the COVID-19 pandemic with survey-based simulations. *PLoS ONE*, 16(10):e0259108, Oct. 2021.
4. S.-J. Kang and S. I. Jung. Age-Related Morbidity and Mortality among Patients with COVID-19. *Infect Chemother*, 52(2):154–164, June 2020.
5. M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. Identification of influential spreaders in complex networks. *Nature Phys*, 6(11):888–893, Nov. 2010.
6. L. Matrajt, J. Eaton, T. Leung, and E. R. Brown. Vaccine optimization for COVID-19: Who to vaccinate first? *Science Advances*, 7(6):eabf1374, Feb. 2021.
7. J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, and W. J. Edmunds. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Medicine*, 5(3):e74, Mar. 2008.
8. L. Pellis, F. Ball, S. Bansal, K. Eames, T. House, V. Isham, and P. Trapman. Eight challenges for network epidemic models. *Epidemics*, 10:58–62, Mar. 2015.
9. F. Pozo-Martin, M. A. Beltran Sanchez, S. A. Müller, V. Diaconu, K. Weil, and C. El Beheraoui. Comparative effectiveness of contact tracing interventions in the context of the COVID-19 pandemic: a systematic review. *Eur J Epidemiol*, 38(3):243–266, Mar. 2023.
10. D. Zhang, Y. Wang, and Z. Zhang. Identifying and quantifying potential super-spreaders in social networks. *Sci Rep*, 9(1):14811, Oct. 2019.