9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Trading contact tracing efficiency for finding patient zero

Keywords: contact tracing, source detection, networks, diffusion, infection

As the COVID-19 disease has demonstrated, identifying the origin of a pandemic remains a challenging task. The search for patient zero may benefit from the widely-used and well-established toolkit of contact tracing methods, although this possibility has not been explored to date. We fill this gap by investigating the prospect of performing the source detection task as part of the contact tracing process, i.e., the possibility of tuning the parameters of the process in order to pinpoint the origin of the infection.

In our experiments, we generate the network structure using either the Barabási-Albert [1], Erdős-Rényi [2], or Watts-Strogatz [5] model. We add a temporal structure of contacts to the network using the generative model by Holme [3]. For a given temporal network, G we randomly select one of the nodes as the source node $v^\dagger$ and start spreading the infection from it. We model the infection using a model of the COVID-19 disease by Rusu et al. [4]. We let the diffusion spread for $T = 28$ rounds, each round corresponding to a single day. Thus, a single simulation run corresponds to four weeks representing the time necessary to discover the disease and assemble a source detection task force. We then perform the contact tracing.

At the beginning of the contact tracing, the set of detected nodes D consists of 10 of the 5% most recently infected nodes (chosen uniformly at random), representing a small set of infections initially noticed by the authorities. We assume a limited budget b indicating the number of people for whom we can perform the contact tracing, the selection of which is based on the value of the *tracing breadth parameter* $\beta_{tr} \in \{1, \ldots, b\}$. Intuitively, increasing the value of β_{tr} makes us focus on the most recently infected nodes, while decreasing the value of β_{tr} makes us reach further into the past. We select β_{tr} members of D that got infected at the earliest time. For every such node v^* we trace their contacts within δ days determined by the *tracing window offset parameter* $\omega_{tr} \in \{0, \ldots, \delta - 1\}$. The contacts of the node v^* are traced on the day $\tau(v^*) + \omega_{tr}$ and the $\delta - 1$ preceding days, where $\tau(v^*)$ is the day when v^* got infected. Intuitively, increasing the value of ω_{tr} shifts the tracing window towards the future, while decreasing the value of ω_{tr} shifts the tracing window towards the past. The probability that v^* remembers the contact that took place at time t is equal to $P(t) = e^{-0.001(T-t)}$ [6]. We test for infection 10 (chosen uniformly at random) of the nodes with which v^* remembered having a contact within the described window). After tracing contacts for all β_{tr} nodes we add the newly identified nodes that were infected to D. We repeat the process until we run out of budget. Unless stated otherwise, in our simulations we consider $\delta = 7$ and $\omega_{tr} = 3$.

Figure 1 presents examples of tracing process for different values of β_{tr} and ω_{tr}. At the beginning of the tracing process, the set of detected nodes consists only of the nodes G and H. If $\beta_{tr} = 2$ then the entire tracing budget $b = 2$ is used to trace contacts of the set of initially detected nodes. On the other hand, if $\beta_{tr} = 1$ then first only the contacts of G are traced (as it was infected earlier than H). As a result, we obtain information about new infected nodes (i.e., node F for $\omega_{tr} = 1$, or nodes F and E for $\omega_{tr} = 0$). We then trace the contacts of the earliest detected infected nodes (i.e., node F for $\omega_{tr} = 1$, or node E for $\omega_{tr} = 0$). As for the ω_{tr} parameter, notice how for $\omega_{tr} = 1$ the contacts are traced at the day of the infection and the subsequent day, while for $\omega_{tr} = 0$ they are traced on the day of the infection and the preceding day. Finally, notice that changing the values of β_{tr} and ω_{tr} can significantly change the outcome of the tracing process, e.g., the actual source of the infection, node A, is detected only for $\beta_{tr} = 1$ and $\omega_{tr} = 0$.

Figure 2 presents the effects of adjusting the breadth parameter β_{tr}. As can be seen, focusing on the breadth of the process results in detecting a greater number of currently infected nodes but a smaller chance of identifying the source of infection, while focusing on the depth of the process makes it easier to identify the source, but decreases the number of currently infected nodes we discover. Let us now more broadly investigate the effects of changing tracing parameters. Figure 3 presents the results of the simulations with varying values of β_{tr} and ω_{tr}. As can be seen, greater values of the tracing breadth parameter β_{tr} result in more effective detection of the currently infected nodes. On the other hand, smaller values of β_{tr} result in closer identifying the source of the infection. These findings confirm our observations based on Figure 2. The value of the ω_{tr} parameter seems to have little effect on the outcome of the tracing process. It might be caused by the fact that contacts between a pair of nodes occur regularly in most cases. Hence, even if we do not notice the contact over which the infection spread, we might observe a different contact between the same two nodes.

Our analysis indicates the existence of a trade-off between identifying the source of the infection process and detecting the nodes that got infected. Figure 4a further investigates this compromise. As it can be seen, getting one step closer to identifying the source (in terms of network distance) costs between 4 and 12 identified infections, depending on the structure of the network, with the cost being the greatest in Watts-Strogatz networks, and the smallest in the Barabási-Albert networks. In Figure 4b we vary the budget b available to the party performing the contact tracing process. Our results suggest that increasing the budget gives a significant advantage in detecting more infected nodes. However, when it comes to identifying the source, the process quickly reaches the state of diminishing returns, where increasing the budget does not provide much better performance. On the other hand, in Figure 4c we vary the time T that the infection process is allowed to run before the contact tracing efforts begin. As can be seen, the percentage of detected infected nodes drops significantly when the infection is allowed to run for a long time. Moreover, identifying the source is getting slightly more difficult as the infection time goes on, although the increase is sublinear with changing T.

The main policy implication of our work comes from the observed trade-off between the effectiveness of contact tracing in terms of finding the source and identifying the infected nodes. Our simulations indicate that, while it is possible to conduct both tasks simultaneously, a performance increase in one of them comes at the cost of a performance decrease in the other. Hence, if the primary goal of the contact tracing process is to curb the spreading of a disease, the emphasis should be on the breadth of the search rather than depth. Such an approach would allow us to maximize the effectiveness of containing the infection while, at the same time, letting us slowly work towards the secondary goal of identifying patient zero. Our results suggest that another key factor is the promptness with which the contact tracing process is started. This finding indicates the need to implement procedures to initiate the process as soon as the first signs of a major infection event are detected. Altogether, a promptly initiated and properly guided contact tracing process can be a crucial tool for combating the next global pandemic.

[1] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 1999.

[2] P. Erdős and A. Rényi. On random graphs I. *Publ. Math. Debrecen*, 1959.

[3] P. Holme. Epidemiologically optimal static networks from temporal network data. *PLoS Comp. Biol.*, 2013.

[4] A. Rusu, K. Farrahi, and R. Emonet. Modelling digital and manual contact tracing for COVID-19 are low uptakes and missed contacts deal-breakers? *medRxiv*, 2021.

[5] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 1998.

[6] P. A. Woźniak, E. J. Gorzelańczyk, and J. A. Murakowski. Two components of long-term memory. *Acta neurobiologiae experimentalis*, 1995.

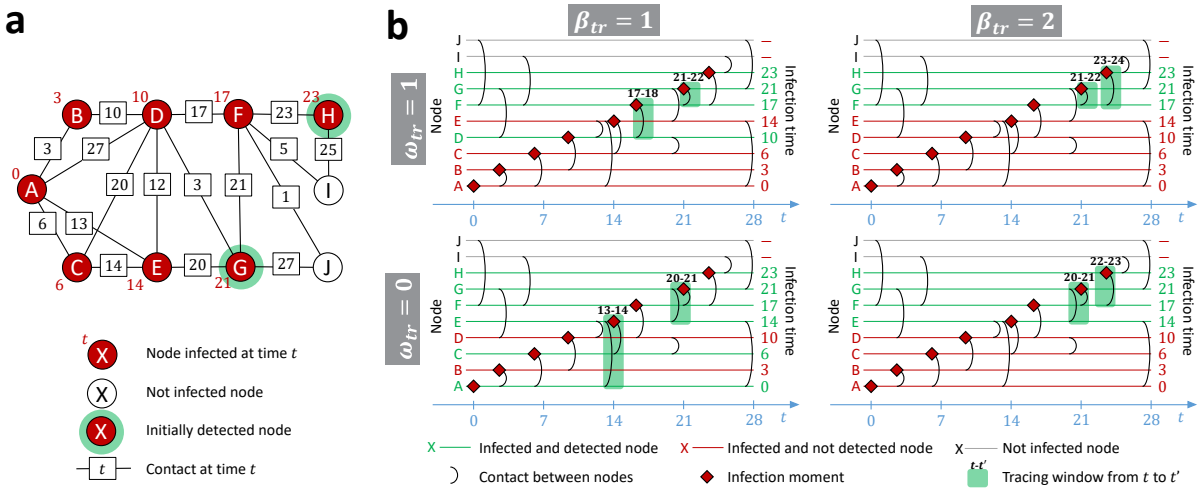

Figure 1: **Examples of the tracing process for different values of tracing breadth β_{tr} and tracing window offset ω_{tr}.** Panel **a** presents a temporal network at the end of the infection process started from the node A. Red numbers next to the nodes represent infection times, white nodes did not get infected. Numbers on the edges represent contact times. At the beginning of the tracing process only the nodes with green rim are detected. The infection process starts from the node A, which is the source (see how the infection time of node A is 0). As a result of the diffusion, all nodes in the network other than the node I and the node J get infected (see how I and J are the only white nodes). Panel **b** presents the results of the tracing process for varying β_{tr} and ω_{tr}, with tracing budget $b = 2$, and tracing window size $\delta = 2$. In every plot horizontal lines represent nodes, with red lines corresponding to infected nodes that did not get detected, green lines to detected infected nodes, and black lines to nodes that did not get infected. Black arcs between lines indicate contacts, with time of the contact represented by the position on the x-axis of the plot. Red rhombuses indicate infection times of the nodes. Green rectangles indicate windows in which contacts of different nodes got traced for the given values of β_{tr} and ω_{tr}. Notice that node A, the actual source of the infection, is detected only for $\beta_{tr} = 1$ and $\omega_{tr} = 0$ (see how the line corresponding to A is green only for this combination of the parameters).

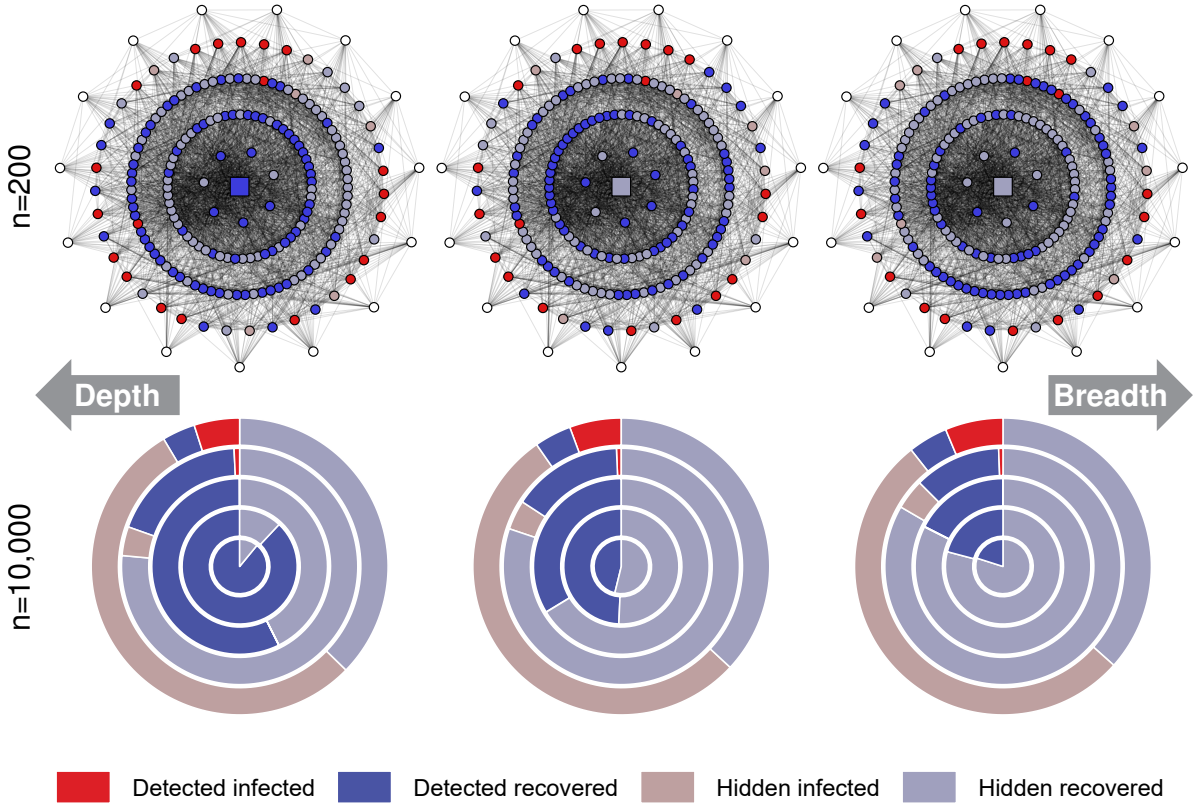

Figure 2: **Tracing process for different values of the breadth parameter β_{tr} in Watts-Strogatz networks.** The first row presents an example of a single simulation of a tracing process in a network with 200 nodes and a budget of $b = 10$. The columns correspond to $\beta_{tr} = 1$, $\beta_{tr} = 5$, and $\beta_{tr} = 10$. The source is represented as the square node in the center. The four inner rings of nodes contain nodes that got infected in the first, second, third, and fourth week, with rings further from the middle corresponding to infections later. Red nodes are infected after 28 days. Blue nodes are recovered. A vivid color indicates that the node got detected by the tracing process for the given value of β_{tr}, and a muted color indicates that the node never got detected (see the legend of the figure). The outer ring contains nodes that never got infected (marked white). The second row presents results for networks with $10,000$ nodes and tracing budget $b = 100$, aggregated over $1,000$ simulation runs. The middle circle represents the diffusion source, while the colored rings represent nodes in the first, second, third, and fourth week, with rings further from the middle corresponding to infections later. The columns correspond to $\beta_{tr} = 1$, $\beta_{tr} = 30$, and $\beta_{tr} = 100$.

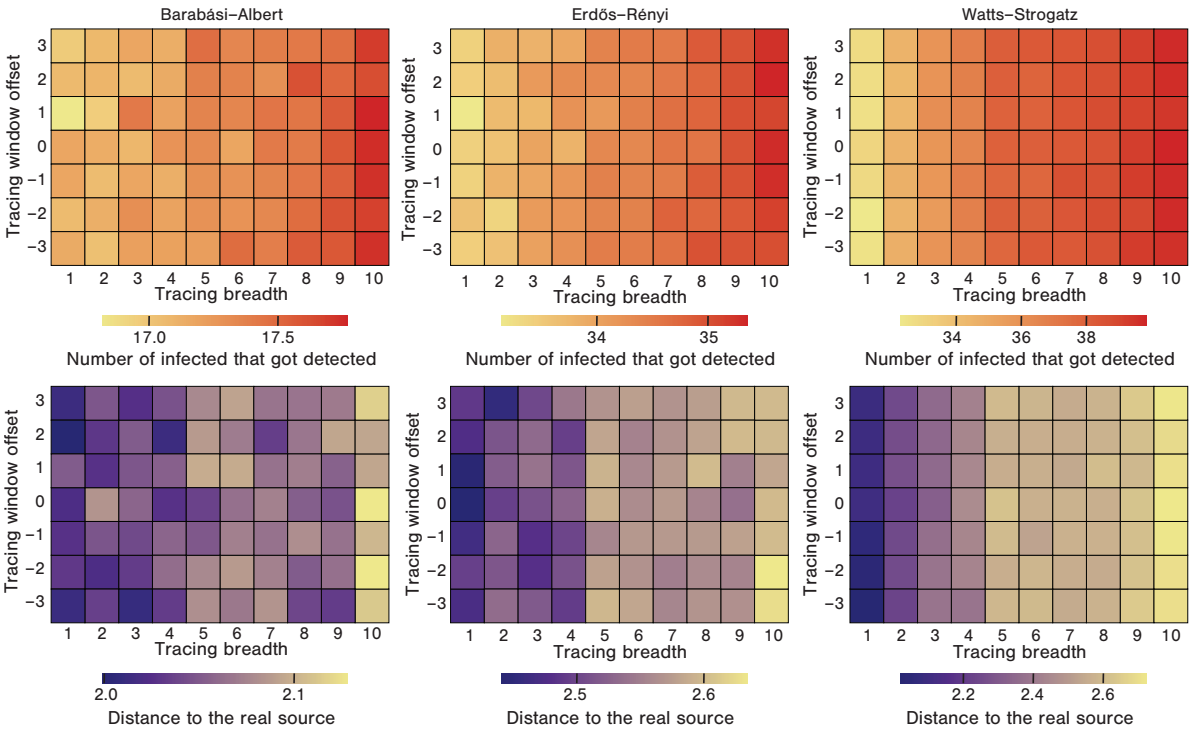

Figure 3: **The effectiveness of tracing for varying β_{tr} and ω_{tr}.** In each plot, the x-axis corresponds to the tracing breadth parameter β_{tr} (with greater values indicating more focus on the breadth). The y-axis corresponds to the tracing window offset parameter ω_{tr} (with greater values indicating the window shifted to the future). The plots in the first row present the number of infected detected by the tracing process, colours closer to red indicate more effective detection. The plots in the second row present the number of edges on the shortest path between the earliest detected infection and the actual source. The colour closer to blue indicates more effective detection. Each column shows results for networks with $10,000$ nodes generated using different models, either Barabási-Albert, Erdős-Rényi, or Watts-Strogatz, with tracing budget $b = 10$. The results are presented as an average of over $1,000$ simulations, with a new network generated for every simulation.

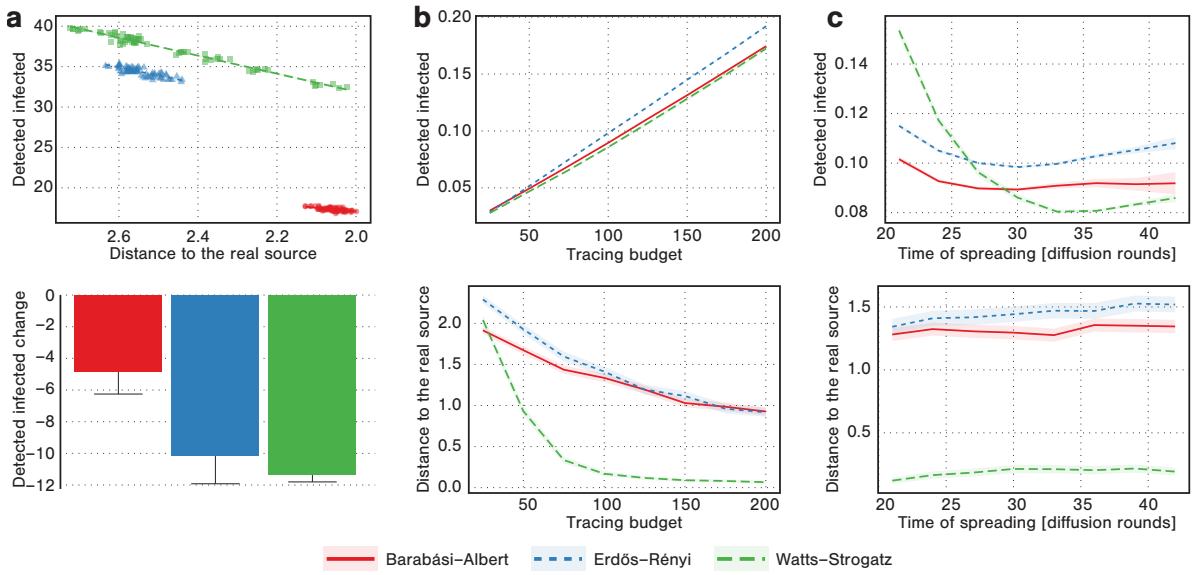

Figure 4: **a. Tracing parametrization.** The scatter plot presents the number of detected infections and the distance to the real source in settings with varying β_{tr} and ω_{tr} parameters (each point is an average value for a specific combination of β_{tr} and ω_{tr}). The lines represent the best fit. The bar plot presents the slopes of the lines, i.e., how much we lose in terms of the number of detected infections when getting one step closer to the source. **b. Changing tracing budget.** In each plot, the x-axis corresponds to the tracing budget b, while the y-axis corresponds to either the number of detected infected nodes, or the number of edges between the earliest detected infection and the real source. **c. Changing infection time.** In each plot, the x-axis corresponds to the infection spreading time T, while the y-axes are the same as in subfigure b. The results in all subfigures are presented as an average of over $1,000$ simulations for networks with $10,000$ nodes, with a new network generated for every simulation using either Barabási-Albert, Erdős-Rényi, or Watts-Strogatz model. The tracing budget is $b = 10$ in subfigure a and $b = 100$ in subfigure c. Error bars and colored areas represent 95% confidence intervals.