Theoretical Models and Preliminary Results
for Contact Tracing and Isolation

Extended Abstract

George Z. Li[1], Arash Haddadan[2], Ann Li[2], Madhav Marathe[2], Aravind Srinivasan[1], Anil Vullikanti[2],
and Zeyu Zhao[1]

[1] University of Maryland, [2] University of Virginia

{gzli929,asriniv1,zachzhao}@umd.edu,
{ahaddada,al7gc,marathe,vsakumar}@virginia.edu

ABSTRACT

Efficient contact tracing and isolation is an effective strategy to control epidemics, as seen in the Ebola epidemic and COVID-19 pandemic. An important consideration in contact tracing is the budget on the number of individuals asked to quarantine—the budget is limited for socioeconomic reasons (e.g., having a limited number of contact tracers). Here, we present a Markov Decision Process (MDP) framework to formulate the problem of using contact tracing to reduce the size of an outbreak while limiting the number of people quarantined. We formulate each step of the MDP as a combinatorial problem, MinExposed, which we demonstrate is NP-Hard. Next, we develop two approximation algorithms, one based on rounding the solutions of a linear program and another (greedy algorithm) based on choosing nodes with a high (weighted) degree. A key feature of the greedy algorithm is that it does not need complete information of the underlying social contact network, making it implementable in practice. Using simulations over realistic networks, we show how the algorithms can help in *bending* the epidemic curve with a limited number of isolated individuals.

KEYWORDS

COVID-19; Contact Tracing; Approximation Algorithms

1 INTRODUCTION

Contact tracing followed by isolation is one of the most effective ways to control epidemics caused by infectious diseases. In this intervention strategy, contact tracers ask infected individuals to report their recent contacts; they then trace these contacts, asking them to isolate for a period of time [5]. Recently, technologies such as the Google-Apple app [1] have provided a solution to augment human contact tracers. When contact tracing apps are used, the strategy is called *digital* contact tracing; otherwise, it is called *manual* contact tracing.

The main limitation of contact tracing is the number of individuals who can be asked to isolate; this number is constrained since isolation imposes a significant economic and social burden to the population. For manual contact tracing, the budget is also dependent on the economic cost of hiring contact tracers. From these constraints, we can see a trade-off between reducing infection spread and minimizing socioeconomic costs. This brings forth a natural question that we study: which individuals should we isolate to make the most effective use of the budget for contact tracing?

Our contributions. We use a Markov Decision Process (MDP) framework to formulate the problem of reducing the size of an outbreak through efficient contact tracing while limiting the number of isolations. At each timestep, the policymaker wants to choose a set of nodes that, when asked to quarantine, minimizes the (expected) number of infections at the end of the epidemic. Since the disease dynamics are constantly changing (due to fluctuating attitudes and behavior), we will only consider finite time horizons of the MDP by solving the combinatorial problem, MinExposed, which focuses on the second neighborhood of the infected set.

- We prove that MinExposed is NP-Hard, so we develop an LP-based algorithm with rigorous approximation guarantees. Using insights from the previous algorithm, we introduce an interpretable and practical greedy approximation algorithm.
- Based on our theoretical results, we devise a heuristic which requires minimal information of the contact graph or disease model, and thus can be made operational in the real world.
- We run simulations of an epidemic with realistic contact networks and parameter values to assess the performance of our heuristic—the results suggest that the heuristic successfully *bends* the epidemic curve.

2 PRELIMINARIES

Let $G = (V, E)$ be the social contact network and let the disease spread on G by an SIR type diffusion process [7]: at each timestep t, an infected node $u \in I(t)$ transmits the disease to each of their neighbors v independently with probability q_{uv}. We assume the currently infected nodes, $I = I(t)$, are known to the policymaker and will self-isolate until they recover. Though all previously infected nodes self-isolate, neighbors of I, which we denote $V_1 = N_G(I) - I$, are already exposed and can continue to spread the disease to the rest of the graph. Policymakers must contact trace these individuals and ask them to isolate. Since this process is expensive and time-intensive for both the government and quarantined individuals, we let B be the budget on the number of nodes that can be reached.

Given this, the objective of policymakers is to minimize the total number of infections in G at the end of the epidemic. This is an idealized problem to solve since the contact graph, transmission rates, and compliance rates are all constantly changing due to various forms of social distancing. As a result, we focus on locally optimal solutions which minimize the expected number of infections in the second neighborhood of I. We denote this neighborhood as $V_2 = N_G(V_1) - I - V_1$ and formalize the problem below.

The MinExposed Problem: Given contact graph $G = (V, E)$, a subset $I \subseteq V$ of infected nodes, transmission probabilities q_{uv} for $(u, v) \in E$, and a budget B, the objective is to find a subset $Q \subseteq V_1$ satisfying $|Q| \leq B$ to quarantine which minimizes the expected number of infections in V_2. In our full paper [6], we also consider the possibility of non-compliance when asked to quarantine.

3 OUR APPROACH

In this section, we present our two algorithms; we defer the proofs of their approximation guarantees and the hardness of MinExposed to our full paper [6]. There, we also show how our methods can be extended to guarantee demographic fairness, both with respect to the people quarantined and the expected number of total infections.

3.1 DepRound

We write MinExposed as a mixed-integer linear program (MILP):

$$\min \sum_{v \in V_2} z_v \qquad \text{s.t.}$$
$$x_u + y_u = 1 \qquad \text{for } u \in V_1$$
$$\sum_{u \in V_1} x_u \leq B \qquad (1)$$
$$z_v \geq y_u \cdot p_u \cdot q_{uv} \qquad \text{for } (u, v) \in E \cap (V_1 \times V_2) \qquad (2)$$
$$x_u, y_u \in \{0, 1\}, z_v \in [0, 1] \qquad \text{for } u \in V_1, v \in V_2$$

We have x_u, y_u for $u \in V_1$ as indicators representing u being quarantined and u potentially spreading the disease, respectively. We allow at most B nodes to be quarantined, as indicated by Constraint 1. Finally, z_v for $v \in V_2$ represents a lower bound on the probability of v getting infected, as conveyed through Constraint 2.

Algorithm 1 DepRound

1: Relax the integer constraints of the MILP to obtain an LP
2: Solve the LP to get vectors $x, y \in [0, 1]^{V_1}$
3: Apply [8] to round vector $x \in [0, 1]^{V_1}$ to get $X \in \{0, 1\}^{V_1}$
4: Output $Q = \{u \in V_1 : X_u = 1\}$

3.2 DegGreedy

In the analysis of DepRound (see [6]), we used the union bound to give an upper bound on the objective value of MinExposed. We now show a simple and interpretable greedy algorithm that directly minimizes the upper bound, and hence gives the same approximation guarantee as DepRound.

Algorithm 2 DegGreedy

1: $w_u \leftarrow p_u \cdot \sum_{v \in V_2, (u,v) \in E} q_{uv}$ for $u \in V_1$
2: pick B nodes with the highest w_u values in V_1 to be in Q

---

4 COMPUTATIONAL EXPERIMENTS

We report on computational experiments conducted to assess the performance of our algorithms. To do this, we use realistic representation of the underlying social network and the disease dynamics.
Disease model. We assume a simple SIR model, with a two-timestep recovery. At each timestep, we have a susceptible set (S), infected set ($I = I_1 \dot\cup I_2$), and a recovered set (R). The partitioning of I models incomplete information: I_1 is the set of recently infected nodes not yet known to be transmitting the disease (due to the incubation period and wait time testing). By the next timestep, I_1 has been tested and becomes I_2, which is known to the policymaker. As a result, all quarantine decisions will be based only on I_2.
Algorithms to compare. In our full paper [6], we devised a practical and implementable variant of DegGreedy, which we called Private DegGreedy. Here, we compare it with two intuitive baselines studied in Armbruster and Brandeau [2]: MostNamed and ListLength. The MostNamed policy selects nodes in V_1 with the most infected neighbors and the ListLength policy is similar, but weighs each neighbor by the inverse of their degree.
Social contact networks. We use a realistic synthetic social contact network for Montgomery County in Virginia constructed by a first principles approach by Barrett et al. [3] and Eubank et al. [4]. For details on the networks and disease model parameters such as transmission probabilities and budget, see our full paper [6].
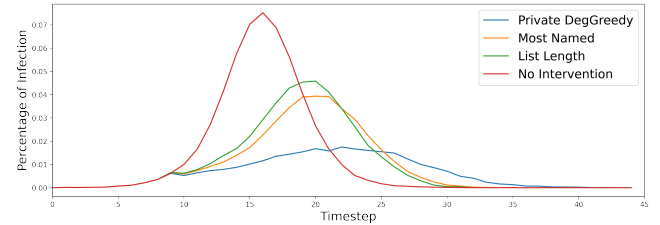

Figure 1: Assessing the performance of the algorithms on realistic instances using agent-based simulations.

Figure 1 shows the effect of the algorithms on the epidemic trajectory. As can be noted the method bends the epidemic curve by (*i*) reducing the total number of infections, (*ii*) delaying the peak, and (*iii*) reducing the size of the peak. The latter two are especially important in practice: a later peak enables time for developing vaccines, which can potentially stop the infection before the peak, and having a smaller peak ensures that people are able to receive adequate treatment in hospitals, which have limited capacity.

REFERENCES

[1] Nadeem Ahmed, Regio A. Michelin, Wanli Xue, Sushmita Ruj, Robert Malaney, Salil S. Kanhere, Aruna Seneviratne, Wen Hu, Helge Janicke, and Sanjay K. Jha. 2020. A Survey of COVID-19 Contact Tracing Apps. *IEEE Access* 8 (2020), 134577–134601. https://doi.org/10.1109/ACCESS.2020.3010226

[2] Benjamin Armbruster and Margaret L Brandeau. 2007. Who do you know? a simulation study of infectious disease control through contact tracing. In *Proceedings of the 2007 Western Multiconference on Computer Simulation*. 79–85.

[3] Christopher L. Barrett, Richard J. Beckman, Maleq Khan, V. S. Anil Kumar, Madhav V. Marathe, Paula E. Stretz, Tridib Dutta, and Bryan Lewis. 2009. Generation and analysis of large synthetic social contact networks. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*. 1003–1014. https://doi.org/10.1109/WSC.2009.5429425

[4] Stephen Eubank, Hasan Guclu, V S Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltán Toroczkai, and Nan Wang. 2004. Modelling disease outbreaks in realistic urban social networks. *Nature* 429, 6988 (May 2004), 180—184. https://doi.org/10.1038/nature02541

[5] Mirjam Kretzschmar, Ganna Rozhnova, and Michiel van Boven. 2020. Isolation and contact tracing can tip the scale to containment of COVID-19 in populations with social distancing. *Available at SSRN 3562458* (2020).

[6] George Li, Arash Haddadan, Ann Li, Madhav Marathe, Aravind Srinivasan, Anil Vullikanti, and Zeyu Zhao. 2021. A Markov Decision Process Framework for Efficient and Implementable Contact Tracing and Isolation. arXiv:2112.15547 [cs.CY]

[7] Madhav Marathe and Anil Vullikanti. 2013. Computational Epidemiology. *Commun. ACM* 56, 7 (2013), 88–96.

[8] Aravind Srinivasan. 2001. Distributions on level-sets with applications to approximation algorithms. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*. 588–597. https://doi.org/10.1109/SFCS.2001.959935