


RESEARCH

Open Access



Scalable and distributed strategies for socially distanced human mobility

Satyaki Roy^{1*}  and Preetam Ghosh²

*Correspondence:

satyakit@unc.edu

¹ University of North Carolina, Chapel Hill, USA

Full list of author information is available at the end of the article

Abstract

COVID-19 is a global health crisis that has caused ripples in every aspect of human life. Amid widespread vaccinations testing, manufacture and distribution efforts, nations still rely on human mobility restrictions to mitigate infection and death tolls. New waves of infection in many nations, indecisiveness on the efficacy of existing vaccinations, and emerging strains of the virus call for intelligent mobility policies that utilize contact pattern and epidemiological data to check contagion. Our earlier work leveraged network science principles to design social distancing optimization approaches that show promise in slowing infection spread however, they prove to be computationally prohibitive and require complete knowledge of the social network. In this work, we present scalable and distributed versions of the optimization approaches based on Markov Chain Monte Carlo Gibbs sampling and grid-based spatial parallelization that tackle both the challenges faced by the optimization strategies. We perform extensive simulation experiments to show the ability of the proposed strategies to meet necessary network science measures and yield performance comparable to the optimal counterpart, while exhibiting significant speed-up. We study the scalability of the proposed strategies as well as their performance in realistic scenarios when a fraction of the population temporarily flouts the location recommendations.

Keywords: Social distancing, Network science, Clustering, Sampling, Parallelization

Introduction

Severe Acute Respiratory Syndrome Coronavirus 2, stated as a global pandemic by the World Health Organization, has infected over 200 million people and led to nearly 4.5 million deaths worldwide. As lockdown and social distancing techniques (<https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-people's-livelihoods-their-health-and-our-food-systems>) became the primary means to combat the soaring infection counts, the impending economic challenges (Overberg et al. 2020) and recent successes in fast-track vaccine development (<https://news.harvard.edu/gazette/story/2020/12/anthony-fauci-offers-a-timeline-for-ending-covid-19-pandemic/>) has encouraged the world leaders to lift lockdown restrictions. However, the late wave of infection surge in several countries, lingering doubts over the effectiveness and health impacts of the vaccines and new virus strains necessitate intelligent public mobility policies that harness contact patterns and epidemiological information

to check the impending threats of contagion during present and future outbreaks (<https://www.healthline.com/health-news/experts-concerned-a-4th-covid-19-wave-may-be-building>).

The lack of foresight and preparedness on the part of the world leaders resulted in the absence of coordinated action plans or public policies. While the world was relying on the findings from the geneticists, doctors and health officials to design make-shift regulations, the epidemiologists, statisticians, and computer scientists explored the socioeconomic and demographic factors contributing to this rapid spread (Adhikari et al. 2020). These efforts included computational and machine learning techniques to predict trends on spread dynamics from epidemiological and clinical data (Wynants 2020; Holmdahl and Buckee 2020; Alimadadi et al. 2020; Randhawa et al. 2020). Their findings lent insights into the epidemiology, causes, clinical manifestation and control measures and helped identify vulnerable communities. Regression analysis and computational approaches were employed as means to gauge effects of testing (Khan et al. 2020; Roy et al. 2021a) and lockdown (Roy and Ghosh 2020) on the pandemic, while unsupervised machine learning and natural language processing approaches broadened our understanding of disease transmissibility and economic challenges (Wang et al. 2020; Roy et al. 2021b).

The accuracy of the parameters of the epidemic models as well as the latter's capability in modeling the epidemiological trends have been key areas of investigation. Holmdahl discussed the constant effort on the part of scientists to refine methods to learn spread dynamics of infectious diseases (Holmdahl and Buckee 2020). Clearly, the predictions from the epidemic models are contingent on factors such as knowledge of demography, infectivity of the virus, accuracy of testing, etc. For instance, Bedi et al. modified the Susceptible-Exposed- Infected-Recovered (SEIR) epidemic model by assuming exposed individuals to be infective and compared the accuracy of their model against that of a Long Short-Term Memory (LSTM) model (Bedi et al. 2020). Gharakhanlou et al. investigated the spread dynamics in Iran by employing agent-based simulation and recommended mitigation measures (Gharakhanlou and Hooshangi 2020), while Ghanam et al. studied the role of government intervention (Ghanam et al. 2020). Furthermore, efforts have been made to analyze the interrelationship between vaccinations, lockdown, mobility and spread (Roy et al. 2021c; Lattanzio and Palumbo 2020) and curb spread through contact-tracing based mobile applications (Kretzschmar et al. 2020; Ferretti et al. 2020; Ahmed et al. 2020; Vax 2014; Nadini et al. 2020; Koppeschaar and Colizza 2017; Dalton et al. 2006). These applications rely on the duration of contact, proximity between individuals and online surveys recording location, patient health and demographic details to identify risk factors. Earlier, we proposed three optimization approaches on social networks that apply network science principles to mitigate contagion by guiding human mobility (Roy et al. 2021d). We carried out simulation experiments using realistic human mobility models and the New York City map to demonstrate that the approaches effectively slow contagion spread. We also designed a mobile application, *MyCovid* (Roy et al. 2021d; Roy 2021) that is presently being deployed to validate the system performance in a realistic setting. However, these approaches face scalability and centralization challenges for large populations.

Contributions

In this work, we present social distancing strategies that optimize the location of individuals residing in urban spaces, such as grocery stores, bus queues, auditoriums, etc., where individuals are likely to engage in social contacts leading to contagion. Our initial efforts in this direction (Roy et al. 2021d) demonstrated three proposed social distancing approaches leveraging network science principles such as *homophily*, *network clustering*, etc. These approaches minimize the number of social ties between the vulnerable (i.e., susceptible individuals) and the vectors of infection (i.e., infected individuals), thereby dampening the rate of infection spread. However, this work is challenged on two fronts: (1) the number of parameters in the optimization scale linearly with the number of individuals, making them computationally prohibitive for large populations and (2) they rely on the knowledge of the entire social network topology. We address these issues in the scalable and distributed social distancing strategies that leverage Markov Chain Monte Carlo Gibbs sampling and grid-based spatial parallelization.

We carry out simulation experiments to show the efficacy of the proposed strategies. We analyze how the system parameters, namely, convergence index and number of grids, can be utilized to tune the optimality vs. scalability trade-off. We gauge the performance of the distributed and sampling strategies in terms of the running time in seconds, optimization score (defined in terms of potential contact between susceptible and infected individuals) as well as the rate of contagion over time in terms of cumulative population of infected, recovered, and dead individuals as per the Susceptible-Exposed-Infected-Recovered-Dead (SEIRD) epidemic model (discussed in “SEIRD Epidemic Model” section). Finally, we show the scalability of the approaches and the effect of a fraction of individuals flouting the recommendations of the system on contagion.

This paper is organized as follows. In “Preliminary Concepts and System Model” section, we discuss the SEIRD epidemic model, preliminary concepts of network science and the system model. In “Approach” section, we present the three social distancing optimization approaches, followed by the scalable and distributed solutions. Sections 4 and 5 deal with the experimental results and discussions. We draw the conclusions in Sect. 6.

Preliminary concepts and system model

Let us first discuss the SEIRD epidemic model, network science concepts (namely, network clustering and homophily) and the system model.

SEIRD epidemic model

The SEIRD model can represent the evolution of the susceptible (S), exposed (E), infected (I), recovered (R) and dead (D) populations (Hethcote 2000). The individuals in S transition to E with rate β , while E transition to I with probability σ ; I transition to R with probability $\gamma \times (1 - \alpha)$ and D with probability $\gamma\alpha$. In other words, γ denotes the proportion of infected that transition to other states, and α is the fraction of those individuals to die. We show the equations corresponding to the state transitions,

where R_0 is the basic reproduction number ranging between 3 and 6 and $\beta = \gamma \times R_0$ (Korolev 2021; Early release-high contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2020).

$$S \rightarrow E \tag{1}$$

$$E \rightarrow I \tag{2}$$

$$I \rightarrow R \tag{3}$$

$$I \rightarrow D \tag{4}$$

Arroyo-Marioli et al. presented an approach to track the rate of contagion in terms of *effective reproduction number* and *growth rate* (defined as rate of increase in daily infection numbers over time) (Arroyo-Marioli et al. 2021). They represented the new infection count at time t in a population of N , as

$$I_t \rightarrow I_{t-1} + \frac{\beta_t \times I_{t-1} \times S_{t-1}}{N} - \gamma I_{t-1} \tag{5}$$

$$\text{Hence, } I_t - I_{t-1} \rightarrow \frac{\beta_t \times I_{t-1} \times S_{t-1}}{N} - \gamma I_{t-1} \tag{6}$$

Given basic reproduction number $R_0 = \beta$ and growth rate $G_t = \frac{I_t - I_{t-1}}{I_{t-1}}$, the effective reproduction number at time t (R_t) is calculated as:

$$R_t = R_0 \times \frac{S_{t-1}}{N} = \frac{\beta_t}{\gamma} \times \frac{S_{t-1}}{N} \tag{7}$$

Plugging R_t in Eq. 6, we get,

$$G_t = \frac{I_t - I_{t-1}}{I_{t-1}} = \gamma \times R_t - \gamma \tag{8}$$

$$R_t = \frac{I_t - I_{t-1}}{I_{t-1}} = 1 + \frac{1}{\gamma} \times G_t \tag{9}$$

Key network science concepts

Given an undirected $G(V, E)$, we discuss the network science concepts that are incorporated by the proposed mobility optimization approaches (details discussed in “[Social distancing optimization](#)” section).

Network clustering

It is the tendency of nodes to form dense communities within G , measured in terms of the number of triads they participate in Holland and Leinhardt (1971), as follows:

$$\zeta(G, u) = \begin{cases} 0 & \text{if } \delta(u) < 2 \\ \frac{2 \times t(u)}{\delta(u) \times (\delta(u) - 1)} & \text{otherwise} \end{cases} \tag{10}$$

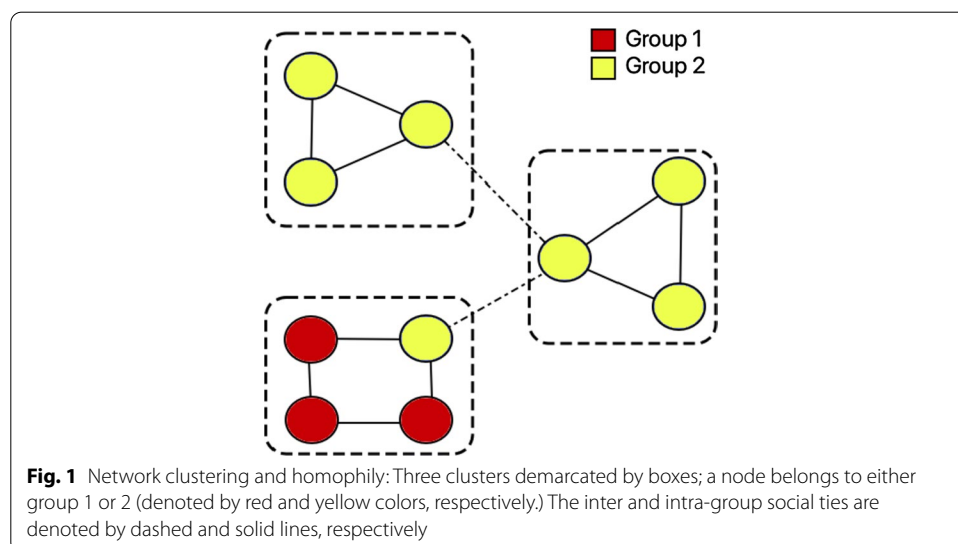
In this equation, $t(u)$ and $\delta(u)$ are the number of triads participated by node u and degree of $u \in V$, respectively. The overall node clustering coefficient of the network (i.e., average clustering of all nodes) shown in Fig. 1, on a scale of 0 and 1, is 0.6.

Homophily

It is the tendency of nodes choosing to attach with nodes of its own *group*, defined as nodes with similar characteristics (Kim and Altmann 2017; McPherson et al. 2001; Kossinets and Watts 2009). We measure homophily in terms of *E-I index*, defined as the difference between proportion of ties between members from different groups and members from the same group (Bojanowski and Corten 2014). An E-I score of -1 means complete homophily, while E-I score of 1 denotes complete heterophily. The network in Fig. 1 has E-I index = -0.6 , making it highly homophilic.

System model

We consider an urban space of dimension $X \times Y$ square feet, where ν mobile individuals are placed. We define *contact threshold* d as the maximum distance between the susceptible and infected individuals such that the susceptible individual may be exposed to the pathogen. We create social network $G_t(V, e_t)$, where V is the set of ν nodes (each representing an individual) and e_t is the set of temporal edges, where an edge $(u, v) \in e_t$ if individuals u and v are within threshold d at time $t \in$ (experiment duration) T . A node, belonging to exactly one epidemic class (S, E, I, R or D) can move within a distance threshold τ of its current location between time t and $t + 1$.



Approach

Each individual carries a smart mobile device capable of communicating with other devices in the vicinity via Wi-Fi or Bluetooth. The neighbor-list of a node u , $n_t(u)$, is the set of individuals that are within distance d at time t . Each individual u must belong to exactly one of S, E, I, R, D states, where $S \cup E \cup I \cup R \cup D = V$. (S, E, I, R and D occasionally have a subscript to denote the number of individuals in that epidemic group at the t -th time epoch.) Fig. 2 shows node 1 relocates to a new location (colored green) within a threshold distance (τ feet) of its old location (colored red).

Social distancing optimization

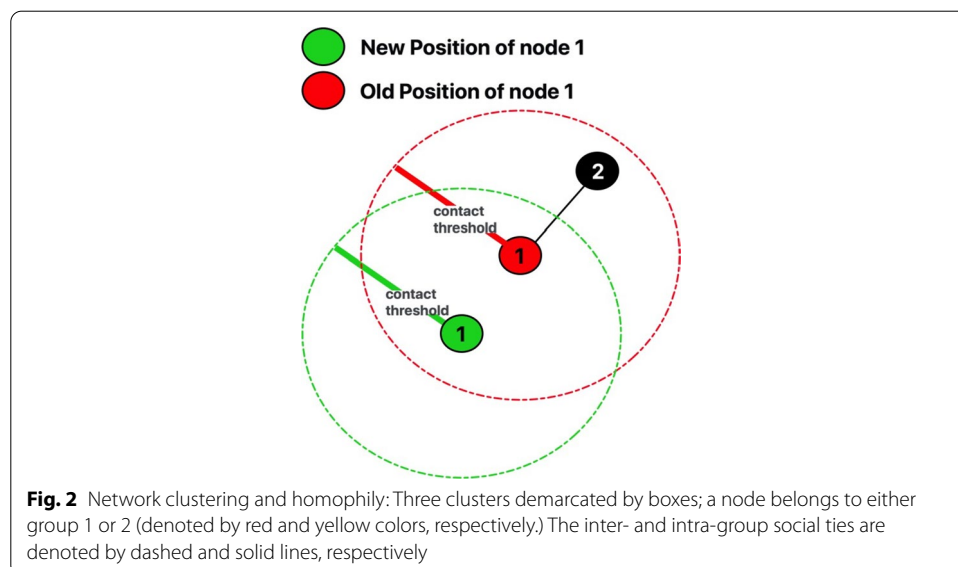
The optimization approaches discussed below were originally proposed in Roy et al. (2021d). Given any social network G_t , these approaches utilize network science principles of network clustering and homophily (see “Key Network Science Concepts” section for details) to generate new locations for individuals (and resultant network G_{t+1}). The goal is to minimize the contact (i.e., links in the social network) between the susceptible and infected individuals and slow the overall contagion.

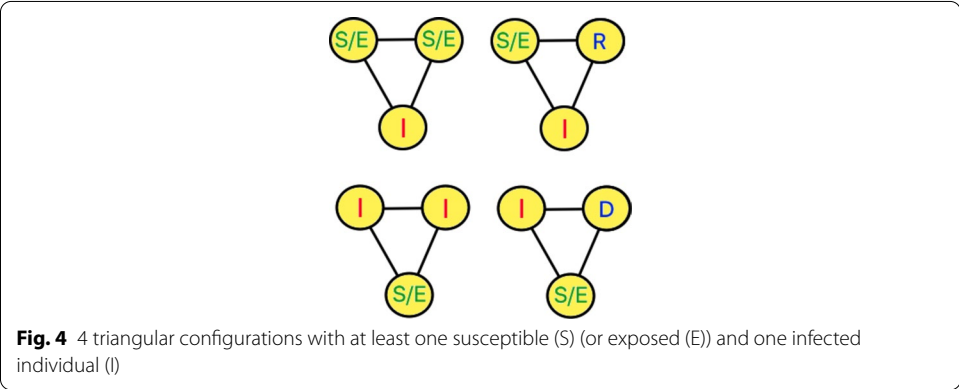
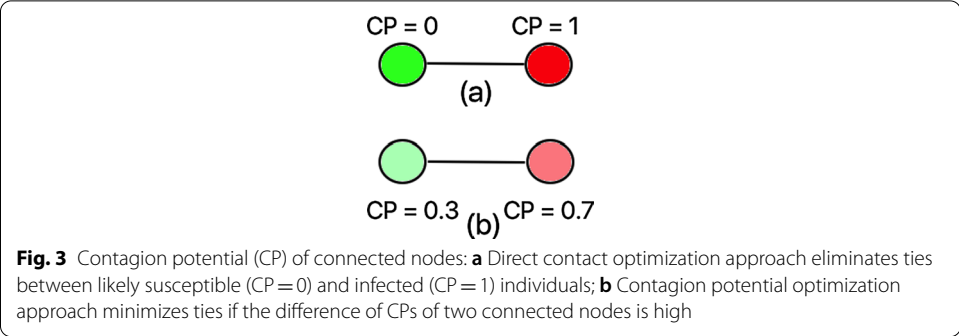
Direct contact approach

It eliminates the contact between the susceptible and infected individuals (see Fig. 3a and Expression 11). This is based on the premise that the infected individuals are the primary sources of contagion, and the susceptible nodes are the target.

Clustering approach

This approach eliminates clusters containing infected individual(s) from the social networks, by re-positioning of nodes (see Expression 12). Recall from “Key Network Science Concepts” section clustering is quantified by the triangle participation of the nodes. Figure 4 shows the four-triangle configurations eliminated by the optimization through node re-positioning.





Contagion potential approach

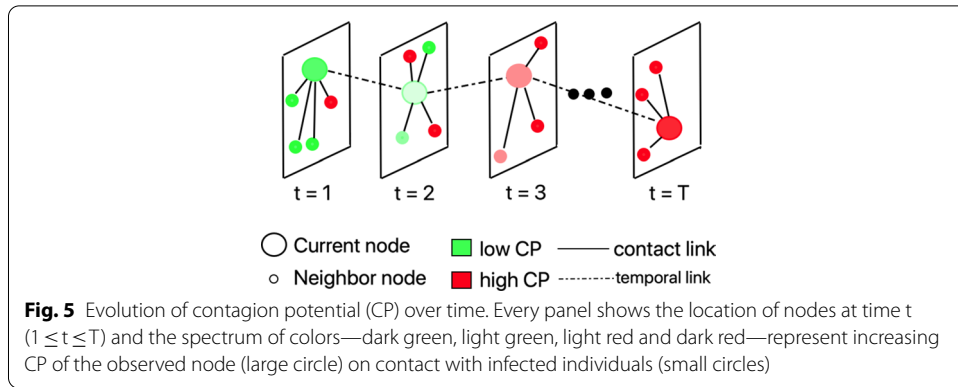
It takes into account the scenario (similar to the model presented in Bedi et al. (2020)) where a person may act as spreader without being tested and identified as infected. We define *contagion potential* (CP) of node u (on a scale of 0 and 1) as its likelihood of acting as spreader. Instantaneous CP is calculated in terms of the number of contacts with individuals with high CP, as follows:

$$\begin{aligned}
 P_t(u) &= 0 && \text{if } t = 0 \\
 P_t(u) &= 1 && \text{if } t \geq 1, u \in I \\
 &= \frac{\sum_{v \in n_t(u)} P_{t-1}(v)}{M_t} && \text{otherwise}
 \end{aligned}$$

M_t is the maximum number of neighbors of any node at time t . Overall CP till time T , Z_T is estimated as the mean over the instantaneous values, as follows:

$$\begin{aligned}
 Z_t(u) &= 0 && \text{if } u \in R, D \\
 Z_t(u) &= 1 && \text{if } u \in I \\
 &= \frac{1}{T} \sum P_t(u) && \text{Otherwise}
 \end{aligned}$$

Figure 5 shows the evolution of the epidemic state of a node over time $t = 1, 2, \dots, T$. The node (depicted as a large circle) has potential of being a spreader due to its lack of contact with other infected individuals. Consequently, it has a low CP (colored green). This approach (formulated in Expression 13) considers the fact that untested



individuals may be infected, and testing can be erroneous. It employs the principle of homophily (refer “[Key Network Science Concepts](#)” section) to group nodes with similar CP into clusters and minimizes the contact between individuals with a high variation in CP. This approach is a generalization of opt-1; instead of representing the infectivity of an individual as a binary case, it assumes a continuous CP value between 0 and 1. In the experimental results discussed in Sect. 4.1, we show that the contagion potential (CP) optimization approach creates more homophilic social networks, by creating links among individuals with similar CP and eliminating links between nodes with dissimilar CPs (see Fig. 3b). This similarity of CP among clustered nodes measured in terms of E-I index minimizes the risk of contact between a potential susceptible and infected individual.

Optimization formulations

In Expression 11, $f(u, v, G_t) = 1$ if nodes u and v are connected (i.e., $(u, v) \in e_t$) in social network G_t , and 0 otherwise. Note that $v \in I_t$, while $u \in S_t$ or E_t because the susceptible and exposed are both asymptomatic and indistinguishable in the real world. Function $\delta(u, v, w, G_t)$ in Expression 12 is equal to 1, if $u, v, w \in V$ form a triangle with at least one infected node, i.e.,

1. $(u, v), (v, w), (u, w) \in e_t$ and
2. $u \in S_t/E_t \mid \mid v \in S_t/E_t \mid \mid w \in S_t/E_t$ and $u \in I_t \mid \mid v \in I_t \mid \mid w \in I_t$

The function $\delta(u, v, w, G_t) = 0$ otherwise.

$$\min_{C_{t+1}} \sum \sum f(u, v, G_t) \tag{11}$$

$$\min_{C_{t+1}} \sum_{u \in S, E, I} \sum_{u \in S, E, I; v > u} \sum_{u \in S, E, I, w > v} \delta(u, v, w, G_t) \tag{12}$$

$$\min \sum_{(u, v) \in E_t; u, v \in S, E, I} |Z_t(u) - Z_t(v)| \tag{13}$$

$$s.t. abs(C_{t+1}(u) - C_t(u)) \leq \tau \tag{14}$$

Expression 13 minimizes the contact between individuals with a high difference in contagion potential (CP), by grouping nodes with similar CP. Given the location of node u at time t , $C_t(u) = (x_t(u), y_t(u))$, Inequality 14 ensures that distance between the current location of any node u at time t , $C_t(u)$, and his location at $t + 1$, $C_{t+1}(u)$ is bounded by the distance threshold τ feet.

Scalable solutions

In the optimization strategies (see “[Social distancing optimization](#)” section), we are looking for the next location of each individual in the social graph G , such that the optimization goals (Eqs. 11 - 14) may be met. Therefore, the optimizer must output vector $C = [C_1, C_2, \dots, C_n]$, where $C(u) = (x(u), y(u))$ is the coordinate of individual u . This raises scalability challenges for large n . To address this, we propose two scalable social distancing strategies, namely the sampling and grid-based strategy. We define *time epoch* as follows:

- *Sampling approach.* similar to MCMC Gibbs Sampling, all nodes attempted to re-locate exactly once within the epoch, assuming all other nodes are fixed
- *Grid-based approach.* one run of the optimization approaches (discussed in “[Social distancing optimization](#)” section) within the grid

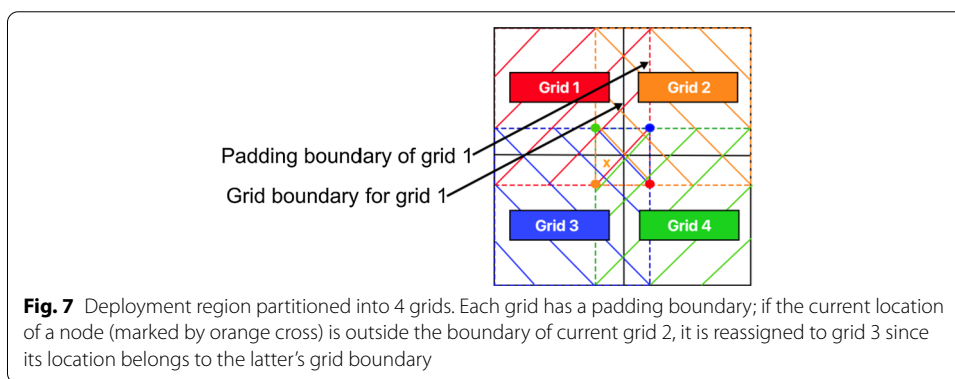
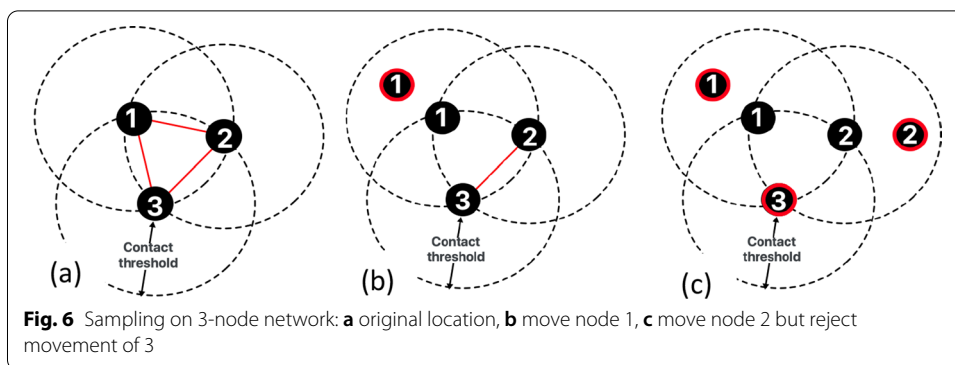
For the direct contact, clustering, and contagion potential approaches, we calculate scores on social network as Expressions 11, 12 and 13, respectively, and the optimization goals are to minimize these scores calculated in all three optimizations based on the (1) social ties between S and I nodes, (2) triangles with S and I nodes and (3) difference in CP of connected nodes, respectively.

Sampling strategy

The sampling strategy is inspired by the Markov Chain Monte Carlo (MCMC) approach, namely Hastings-Metropolis (Carlo 2004). In each time epoch t and social network G_t , we iteratively sample a node u at a time with equal probability and attempt to re-locate it if other nodes $V(G_t) \setminus u$ do not move. The optimizer is invoked to place u at locations within radius τ of current location $C_t(u)$. The move is accepted if the resultant social network minimizes scores and rejected otherwise. The time epoch is complete when the convergence criteria, the fraction of total relocated nodes $VR(G_t)$ is less than a threshold π , i.e., $\frac{VR(G_t)}{V(G)} < \pi$, is satisfied. In Fig. 6, we demonstrate the above steps for a 3-node social network.

Grid-based strategy

We parallelize the optimization approaches by partitioning the deployment region into grids. Figure 7 depicts a scenario where the region is partitioned into 4 grids colored red, orange, blue and green. At each time epoch, every grid with the nodes placed within it at the time is initialized in a parallel process. Each grid has a padding region, whose area extends from the horizontal and vertical border of each grid, by length equal to the distance threshold τ . Since any node can move a distance of τ in each epoch, a node



belonging to a grid experiences an illusion that it is not restricted by the grid boundary and can reside anywhere within the padded boundary (see Fig. 7).

Approach We consider a master–slave paradigm running a map reduce approach. Given Z spatial grids, in a time epoch, the master *maps* grids to parallel slave processes and the optimization approaches are simultaneously and independently invoked in the different grids. At the end of an optimization run, the processes return the optimized locations from their grids to the master. The master performs the *reduce* step where any node u , with location $C(u)$, located (outside its grid boundary and) in padded region is assigned to another grid if $C(u)$ belongs to the grid's boundary. Consider the location of a node, marked in orange cross, originally belonging to grid 2 is reassigned to grid 3 as its new location is within the latter's grid boundary.

Hybrid strategy

We combine the sampling- and grid-based approach in order to achieve greater scalability. We follow three steps: (1) the deployment region is partitioned into grids, (2) within each grid the sampling strategies are invoked by master, and (3) slave processes return the optimized locations of its nodes once the sampling convergence criterion is achieved. The above steps are repeated in each time epoch.

Observations Note that high convergence index or grid count in the sampling and grid-based approaches result in greater speed-up at the expense of optimality of the optimi-

zation goal (as shown in the experimental results in “Results” section). In the rest of the paper, we use the term *approach* to refer to the three optimizations (“Social distancing optimization” section) and the term *strategy* to refer to sampling- (or distributed) and grid-based solutions.

Results

We create a platform on Python SimPy discrete event simulation environment (Matloff 2009), where each node is an agent, and the total time is divided into discrete time epochs. The simulation environment enforces the differential equations of the SEIRD epidemic model (Eqs. 1–4) indirectly as follows: the social network at any time epoch is implemented through a spatial model, where moving agents are nodes that form a temporal social tie when they are within contact threshold d . This section has the following subsections: (1) optimization versus sampling approaches, (2) effect of convergence index, the performance of (3) distributed strategy, and (4) grid-based solution and (5) scalability analysis (Table 1).

Default parameters

We carry out experiments on 2.6 GHz 6-Core Intel Core i7 macOS 16 GB RAM, each of duration 100 time epochs, on a population ranging from 15–4000 individuals and contact rate $\beta=0.55$. We plot mean curve from 25 iterations, showing the cumulative count, which we measure as the sum of infected, recovered, and dead individuals at a given time. To ensure fairness of comparison, individuals have the same initial starting location and epidemic status in each run of the experiment. The contact threshold is $d = 6$ ft. and individuals move within distance threshold $\tau = 25$ ft. on an average at every time epoch. All three strategies are run using the SEIRD model for contagion spread. We compare the scalable and distributed solutions against the random mobility strategy. Grid-based parallelization was achieved using the Python Multiprocessing library (Palach 2014). It is worth repeating that we define scores for the scalable versions of the three optimization approaches in terms of the values of expressions 11, 12 and 13; the lower the score, the closer the scalable solution is to the optimal solution (Table 2).

Table 1 Default parameter values

Parameter	Notation	Value
Number of iterations	-	25
Population size	v	50–4000
Simulation area length	X	50–400 ft.,
Simulation area breadth	Y	50–400 ft
Simulation duration	T	50, 100 time epochs
SEIRD parameters (Korolev 2021)	$\sigma, \gamma, \alpha, \beta$	0.25, 0.1, 0.05, 0.55
Contact threshold	d	6 ft
Distance threshold	τ	25 ft
Init. sus. and inf. fraction	-	0.7, 0.3
Convergence index	π	0.3–0.5
Grid count	Z	1–64

Table 2 No. of nodes vs. area

Nodes	75	90	105	120	135	150
Area (sq. ft)	100 × 50	100 × 60	100 × 70	100 × 80	100 × 90	100 × 100

Optimization versus sampling approaches

We calculate the running time and scores of direct contact and clustering approaches against the corresponding sampling approaches with convergence index 0.3. Figure 8a shows that sampling strategy for direct contact approach is much closer to its optimization counterpart than clustering approach. With respect to the running time in seconds, the optimization approaches exhibit a significantly higher growth rate than the sampling versions (see Fig. 8b). We apply nonlinear curve-fitting to fit the running time to polynomials of order 2 (i.e., $y = c_0 + c_1x_1 + c_2x_2$). In Table 3, we show that the direct contact and clustering approaches have higher coefficient or order 2 (c_2), resulting in higher running time than the sampling counterparts.

Recall from “Key Network Science Concepts” section, homophily of a network is measured in terms of E-I index. Since optimization contagion potential approach (Expression 13) attempts to achieve homophily by grouping nodes with similar CPs. We compare the E-I indices of the original and (sampling approach) modified networks. We discretize the node CPs by rounding them off to one decimal place (i.e., 0.05 becomes 0.1) and record the E-I indices. Figure 8c shows that the E-I indices of the optimized networks are significantly lower (even negative), suggesting that they exhibit a higher proportion of links among nodes with similar CPs.

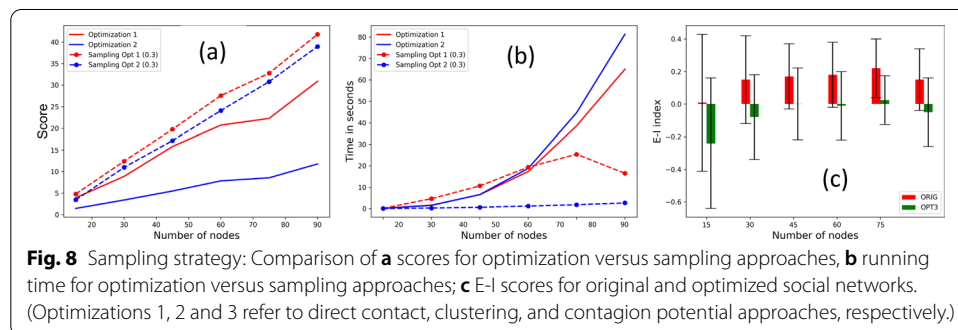


Table 3 Polynomial coefficients from the nonlinear curve fitting (of order 2) on the running time in seconds against the number of nodes, for approach 1 (i.e., direct contact) and approach 2 (i.e., clustering)

Approach	Coeff. 0 (c_0)	Coeff. 1 (c_1)	Coeff. 2 (c_2)
Opt-1 (Direct contact)	0.98	-4.24	3.39
Sam-1 (Direct contact)	-2.62	11.70	-1.59
Opt-2 (Clustering)	1.82	-7.54	4.63
Sam-2 (Clustering)	0.04	0.08	0.07

Effect of Convergence Index

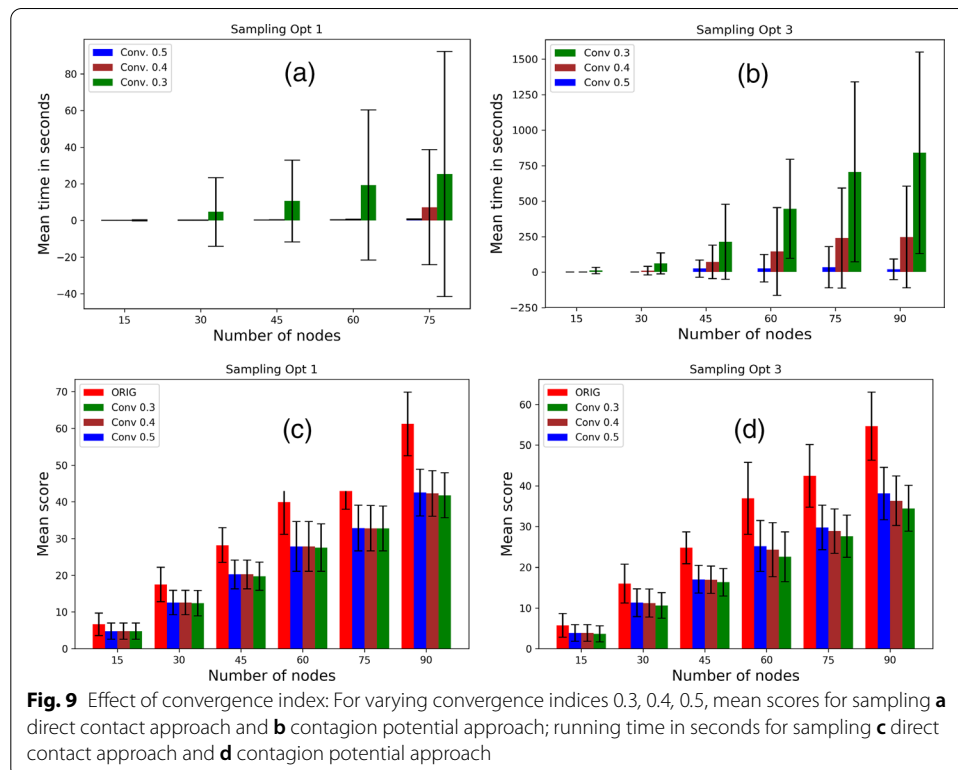
We vary the convergence indices—0.3, 0.4, 0.5—and record the running scores and running time for sampling approaches 1 and 3. We study the trade-off offered by the convergence parameter. Figure 9a, b show that starting with the original scores (colored red), the sampling approaches exhibit better scores. For both approaches, lowering in convergence does not cause a major improvement. Figure 9c, d show that, for both approaches, convergence indices 0.4 and 0.5 greatly outperform 0.3 in terms of running time.

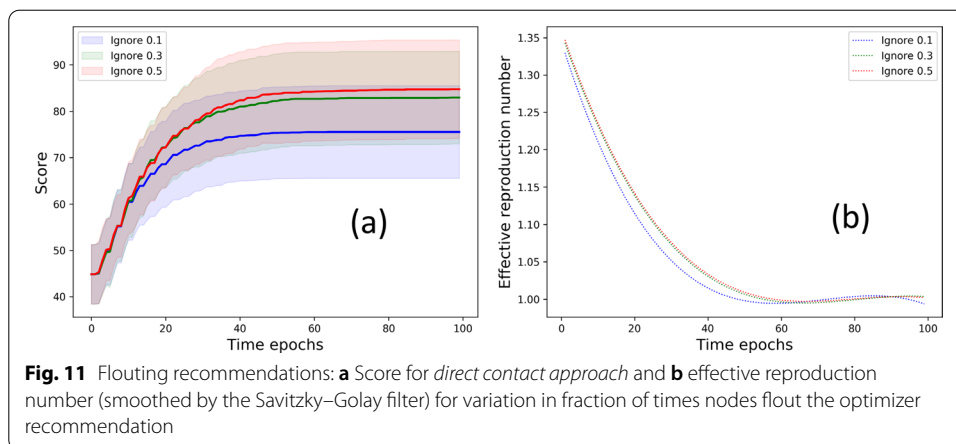
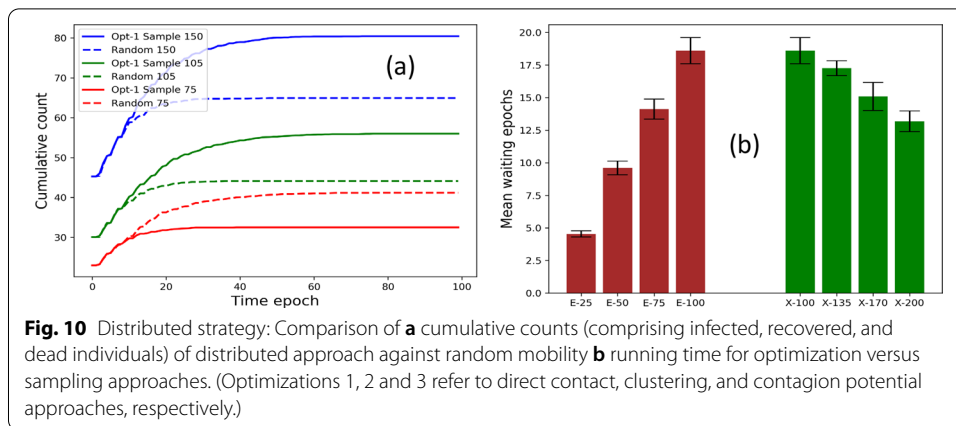
Performance of distributed strategy

We estimate the performance of the distributed approach (defined in “*Scalable solutions*” section) with respect to the cumulative count (constituting infected, recovered, and dead population) and waiting times needed to wait for their neighbors with lower IDs to move. Figure 10a shows that for 75, 105, 100 nodes, the distributed approach exhibits a lower cumulative count (i.e., slower contagion) over time.

We also consider the two variables that may result in variable waiting times – number of nodes and deployment area. Figure 10b shows that (1) varying population 25, 50, 75, 100 for fixed area 100 × 100 sq. ft. causes a linear growth in mean waiting epochs per node (brown bars); similarly, (2) for a fixed population of 100 individuals and varying the area of 100 × 100, 135 × 100, 170 × 100 and 200 × 100 results in a decrease in the mean waiting time epochs per node.

Flouting recommendation. We study the effect of flouting the location recommendation of the social distancing strategy. We record the score when the nodes follow sampling approach 1 for 90%, 70%, 30% of time and thereby undertake random

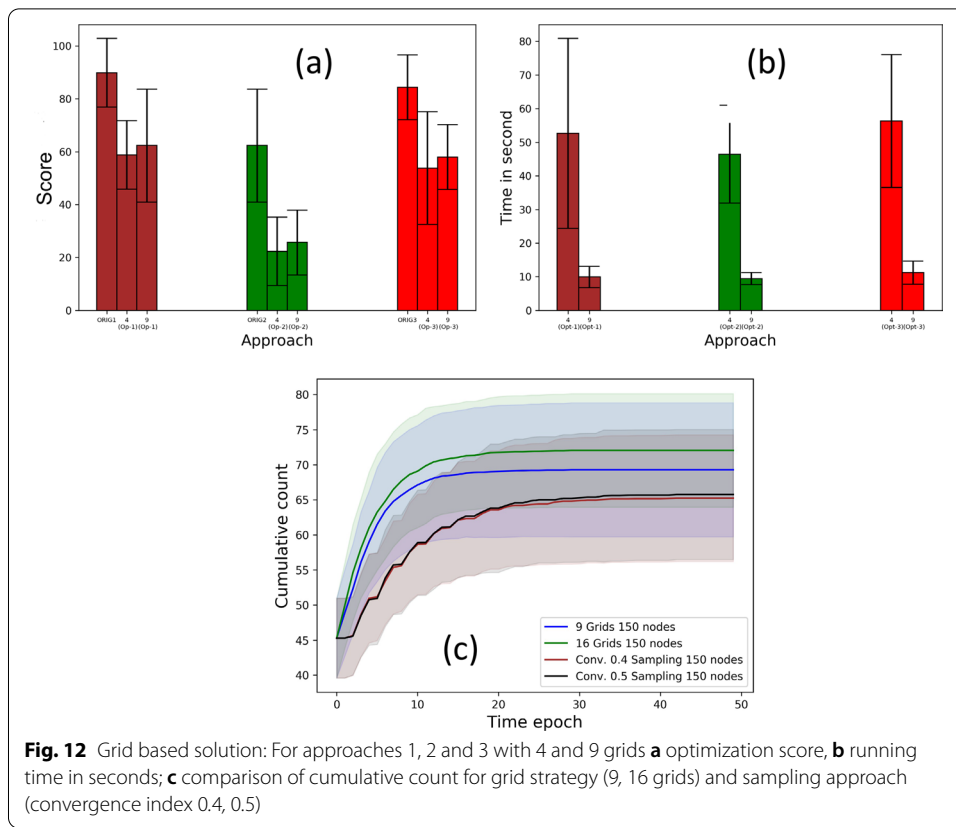




mobility 10%, 30%, 50% at other times. Figure 11a shows that scores are hampered as the nodes increasingly ignore recommendation. This result may also be viewed in light of situations where certain nodes may temporarily get discharged or fall off the grid and lose contact with their immediate neighbors. Furthermore, we plot the effective reproduction number (refer to “SEIRD Epidemic Model” section for details) that provides a realistic measure of the number of secondary cases caused per infected individual. Figure 11b shows that the effective reproduction number (smoothed by the Savitzky–Golay filter (Press and Teukolsky 1990)) is the least when the population obeys the mobility recommendations.

Performance of grid-based solution

The goal of the grid-based solution is to achieve a running time vs. performance trade-off. We evaluate the efficacy of the grid-based approach with respect to the score and running time in seconds for varying grid counts. For a social network of 150 nodes, Fig. 12a shows the mean original score and the improved scores with 4 and 9 grids with direct contact approach; the scores achieved by the two grid configurations are comparable. With respect to the running time in seconds, Fig. 12b shows



that the 9-grid configuration yields approximately 5 times speed up compared to the 4-grid counterpart, proving the efficacy of the parallel solution.

We compare the performance of the sampling and grid-based solutions with respect to the cumulative counts. Recall from our discussion in “[Sampling strategy](#) and [Grid-based strategy](#)” section that a (1) high convergence index in the sampling approach or (2) large number of grids in the grid-based approaches, yield higher speed-up at the cost of the optimality of the optimization objectives. Figure 12c shows that the sampling strategy with convergence indices 0.4, 0.5 achieve a significantly lower mean count than 4- and 9-grid configurations for networks of 25 nodes.

Scalability analysis

We analyze the improvement achieved in running time due to the grid and sampling approaches by recording the running time for 250, 500, 750, 1000 nodes with area $200 \times (1) 50, (2) 100, (3) 150$ and $(4) 200$ sq. ft., respectively. For the sampling and grid strategies we use convergence index 0.4 and 16 grids, respectively. Figure 13a shows that, for direct contact approach, the sampling strategy scales better than the grid-based strategy.

We also evaluate how the running time compares for a hybrid of sampling and grid strategies on direct contact approach, particularly when more computational resources are employed. Recall from “[Hybrid strategy](#)” section, the hybrid approach uses grid-based parallelization and the sampling approach with the grids. We consider four

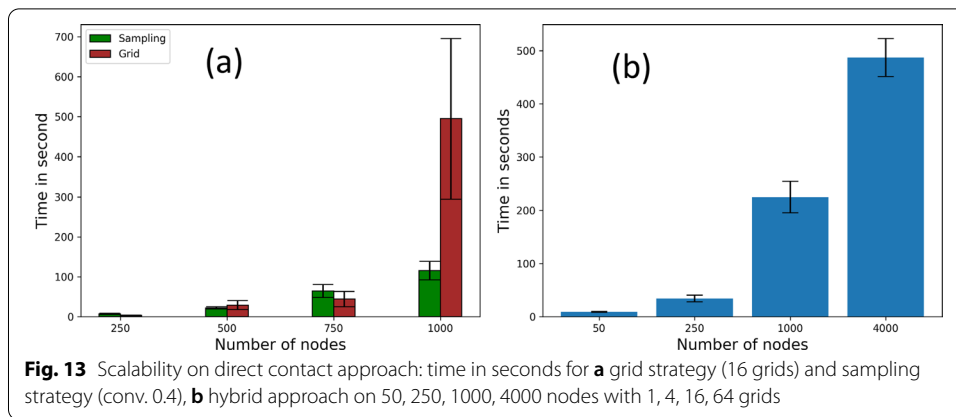


Table 4 No. of nodes vs. area

Nodes	50	250	1000	4000
Area (sq. ft)	50 × 50	100 × 100	200 × 200	400 × 400
Grids	1	4	16	64

settings (summarized in Table 4). Figure 13b shows that the running time for the hybrid strategy grows proportionally with the order of the social network. The growth rate in this experiment, with increasing computational resources, is lower than that reported in Fig. 13a where the number of grids is kept constant.

Discussions

Our simulations suggest that the proposed strategies mitigate the scalability challenges of solving the three optimizations for large populations. In addition to the speed up exhibited by the sampling, grid-based and hybrid strategies, the distributed algorithm enables each node to operate solely on the knowledge of immediate neighbors, as opposed to the entire social network topology. It however raises a few questions and offers new research directions. First, the dynamics of mobility in an urban setting is highly noisy (characterized by (1) erratic movements and (2) uneven spatial population density), making it imperative to deploy the system in a real setting to study their running time complexity and load balancing. To achieve this, we have designed a MyCovid mobile application (Roy et al. 2021d; Roy 2021) that is currently being used by a small population of students to validate the original optimization strategies. Second, although we have tested the optimization on human mobility models, we will need to incorporate the fact that individuals may have predetermined source and destination locations that may override the recommendations of the optimization strategies. This requires an online algorithm to learn personalized schedules and itineraries to make informed recommendations. Similarly, we shall compare the performance of the proposed strategies for small and large population sizes. For accurate predictions with small population, we shall incorporate the necessary correction factors (Grima 2010). Third, there are important security and privacy considerations associated with location sharing. Although the distributed strategy annuls the need to know the entire social network topology,

individuals may exhibit reluctance to be detected by neighbor devices, making it essential to build adaptive models that can work with uncertainty as well as infuse identity detection and privacy-preserving techniques into the system. Fourth, it is worth exploring dynamic algorithms that autonomously adjust system parameters like convergence index and grid count (or size) based on the influx or outflow of nodes in urban space.

Conclusion

In this paper, we presented scalable and distributed social distancing strategies to inform the mobility of individuals roaming in an urban space. The proposed strategies leverage network science principles, such as homophily and network clustering, in conjunction with MCMC Gibbs random sampling and grid-based spatial parallelization. In addition to scaling well for large social networks, the distributed strategy allows individuals to determine next locations without knowledge of the entire network topology. We perform simulation experiments to delineate how one can tune system parameters such as convergence index and grid count to achieve trade-off between running time and rate of contagion. We compare the performance of the proposed strategies, as well as their hybrid, against random human mobility for varying human population sizes and analyze how ignoring optimization recommendations affect overall infection spread.

Abbreviations

LSTM: Long Short-Term Memory; S: Susceptible; E: Exposed; I: Infected; R: Recovered; D: Death; CP: Contagion potential; MCMC: Markov Chain Monte Carlo.

Acknowledgements

We acknowledge the Department of Computer Science, Virginia Commonwealth University for its computational resources.

Authors' contributions

SR and PG conceived of the idea presented in this paper. SR developed the theory and performed the necessary experiments. SR and PG verified the methods and results. All authors discussed the results and contributed to the final manuscript. Both authors read and approved the final manuscript.

Funding

This work is partially supported by National Science Foundation (CBET- 1802588).

Availability of data and materials

All relevant data (epidemiological and demographic data related to the boroughs of New York City) as well as the Python scripts are made available at <https://github.com/satunr/COVID-19/tree/master/Network%20Science/> ScriptScalable.

Declarations

Competing interests

The authors have declared that no competing interests exist.

Author details

¹University of North Carolina, Chapel Hill, USA. ²Virginia Commonwealth University, Richmond, USA.

Received: 10 June 2021 Accepted: 30 October 2021

Published online: 13 December 2021

References

- Adhikari S, Meng S, Wu Y, Mao Y, Ye R, Wang Q, Sun C, Sylvia S, Rozelle S, Raat H et al (2020) Epidemiology, causes, clinical manifestation and diagnosis, prevention, and control of coronavirus disease (covid-19) during the early outbreak period: a scoping review. *Infect Dis Poverty* 9(1):29. <https://doi.org/10.1186/s40249-020-00646-x>
- Ahmed N et al (2020) A survey of covid-19 contact-tracing apps. *IEEE Access* 8:134577–134601. <https://doi.org/10.1109/ACCESS.2020.3010226>
- Alimadadi A, Aryal S, Manandhar I, Munroe P, Joe B, Cheng X (2020) Artificial intelligence and machine learning to fight covid-19. *Physiol Genomics* 52(4):200–202. <https://doi.org/10.1152/physiolgenomics.00029.2020>

- Arroyo-Marioli F et al (2021) Tracking r of covid-19: a new real-time estimation using the kalman filter. *PLoS ONE* 16(1):e0244474
- Bedi P, Gole P, Gupta N, Jindal V et al (2020) Projections for covid-19 spread in India and its worst affected five states using the modified seird and Istm models. arXiv preprint. Available: [arXiv:2009.06457](https://arxiv.org/abs/2009.06457)
- Bojanowski M, Corten R (2014) Measuring segregation in social networks. *Soc Netw* 39:14–32. <https://doi.org/10.1016/j.socnet.2014.04.001>
- Carlo C (2004) Markov chain Monte Carlo and gibbs sampling. Lecture notes for EEB, 581
- Dalton C, Durrheim D et al (2009) Flutracking: a weekly Australian community online survey of influenza-like illness in 2006, 2007 and 2008. *Commun Dis Intell Quart Rep* 33(3):316–322
- Experts concerned a 4th covid-19 wave may be building. <https://www.healthline.com/health-news/experts-concerned-a-4th-covid-19-wave-may-be-building>, 2021
- Fauci says herd immunity possible by fall, 'normality' by end of 2021. <https://news.harvard.edu/gazette/story/2020/12/anthony-fauci-offers-a-timeline-for-ending-covid-19-pandemic/>, 2021
- Ferretti L et al (2020) Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science*. <https://doi.org/10.1126/science.abb6936>
- Ghanam R, Boone E, Abdel-Salam A (2020) Seird model for Qatar covid-19 outbreak: a case study. arXiv preprint. [arXiv:2005.12777](https://arxiv.org/abs/2005.12777)
- Gharakhanlou N, Hooshangi N (2020) Spatio-temporal simulation of the novel coronavirus (covid-19) outbreak using the agent-based modeling approach (case study: Urmia, Iran). *Informatics in Medicine Unlocked* 20:100403. <https://doi.org/10.1016/j.imu.2020.100403>
- Grima R (2010) An effective rate equation approach to reaction kinetics in small volumes: theory and application to biochemical reactions in nonequilibrium steady-state conditions. *J Chem Phys* 133(3):07B604
- Hethcote H (2000) The mathematics of infectious diseases. *SIAM Rev* 42(4):599–653. <https://doi.org/10.1137/S0036144500371907>
- Holland P, Leinhardt S (1971) Transitivity in structural models of small groups. *Comp Group Stud* 2(2):107–124. <https://doi.org/10.1177/104649647100200201>
- Holmdahl I, Buckee C (2020) Wrong but useful—what covid-19 epidemiologic models can and cannot tell us. *N Engl J Med* 383(4):303–305. <https://doi.org/10.1056/NEJMp2016822>
- Khan N, Naushad M, Fahad S, Faisal S, Muhammad A (2020) Covid-2019 and world economy. *J Health Econ*. <https://doi.org/10.2139/ssrn.3566632>
- Kim K, Altmann J (2017) Effect of homophily on network formation. *Commun Nonlinear Sci Numer Simul* 44:482–494. <https://doi.org/10.1016/j.cnsns.2016.08.011>
- Koppeschaar C, Colizza V et al (2017) Influenzanet: citizens among 10 countries collaborating to monitor influenza in europe. *JMIR Public Health Surveill* 3(3):e66. <https://doi.org/10.2196/publichealth.7429>
- Korolev I (2021) Identification and estimation of the seird epidemic model for covid-19. *J Econ* 220(1):63–65. <https://doi.org/10.1016/j.jeconom.2020.07.038>
- Kossinets G, Watts D (2009) Origins of homophily in an evolving social network. *Am J Sociol* 115(2):405–450
- Kretzschmar M et al (2020) Impact of delays on effectiveness of contact tracing strategies for covid-19: a modelling study. *Lancet Public Health* 5(8):e452–e459. [https://doi.org/10.1016/S2468-2667\(20\)30157-2](https://doi.org/10.1016/S2468-2667(20)30157-2)
- Lattanzio S, Palumbo D (2020) Lifting restrictions with changing mobility and the importance of soft containment measures: a seird model of covid-19 dynamics. COVID-19 Economic Research – University of Cambridge. <http://covid.econ.cam.ac.uk/lattanzio-palumbo-importance-of-soft-containment-measures>
- Matloff N (2008) Introduction to discrete-event simulation and the simpy language. Davis, CA. Dept of Computer Science. University of California at Davis. Retrieved on August, 2(2009):1–33
- McPherson M, Smith-Lovin L, Cook J (2001) Birds of a feather: homophily in social networks. *Ann Rev Sociol* 27(1):415–444
- Nadini M, Richmond S, Huang J, Rizzo A, Porfiri M (2020) Design and feasibility study of the mobile application stop the spread. *IEEE Access* 8:172105–172122. <https://doi.org/10.1109/ACCESS.2020.3022740>
- Overberg P, Kamp J, Michaels D (2020) The covid-19 death toll is even worse than it looks. <https://www.wsj.com/articles/the-covid-19-death-toll-is-even-worse-than-it-looks-11610636840>
- Palach J (2014) Parallel programming with Python. Packt Publishing Ltd
- Press W, Teukolsky S (1990) Savitzky-golay smoothing filters. *Comput Phys* 4(6):669–672
- Randhawa G, Soltysiak M, El Roz H, de Souza C, Hill K, Kari L (2020) Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: Covid-19 case study. *PLoS ONE* 15(4):e0232391. <https://doi.org/10.1371/journal.pone.0232391>
- Roy S (2021) GitHub Repository - MyCovid App. <https://github.com/satunr/COVID-19/tree/master/Network%20Science>
- Roy S, Ghosh P (2020) Factors affecting covid-19 infected and death rates inform lockdown-related policymaking. *PLoS ONE* 15(10):e0241165. <https://doi.org/10.1371/journal.pone.0241165>
- Roy S, Dutta R, Ghosh P (2021b) Recreational and philanthropic sectors are the worst-hit us industries in the covid-19 after-math. *Soc Sci Humanities Open* 3(1):100098. <https://doi.org/10.1016/j.ssaoh.2020.100098>
- Roy S, Dutta R, Ghosh P (2021c) Optimal time-varying vaccine allocation amid pandemics with uncertain immunity ratios. *IEEE Access* 9:15110–15121
- Roy S, Cherevko A, Chakraborty S, Ghosh N, Ghosh P (2021d) Leveraging network science for social distancing to curb pandemic spread. *IEEE Access* 9:26196–26207
- Roy S, Biswas P, Ghosh P (2021) Quantifying mobility and mixing propensity in the spatiotemporal context of a pandemic spread. *IEEE Transactions on Emerging Topics in Computational Intelligence*, pp 1–11
- Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R (2020) Early release-high contagiousness and rapid spread of severe acute respiratory syndrome coronavirus. *PLoS ONE* 15(7):1470–1477. <https://doi.org/10.1371/journal.pone.020282>
- Vax (2014) <https://github.com/digitalepidemiologylab/VaxGame>

- Wang P, Zheng X, Li J, Zhu B (2020) Prediction of epidemic trends in covid-19 with logistic model and machine learning technics. *Chaos Solitons Fractals* 139:110058. <https://doi.org/10.1016/j.chaos.2020.110058>
- World health organization-impact of covid-19 on people's livelihoods, their health, and our food systems. <https://www.who.int/news/item/13-10-2020-impact-of-covid-19-on-peoples-livelihoods-their-health-and-our-food-systems>, 2020
- Wynants L et al (2020) Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ*. <https://doi.org/10.1136/bmj.m1328>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
