Promoting Fair Vaccination Strategies through Influence Maximization: A Case Study on COVID-19 Spread

Nicola Neophytou[1,2]**, Afaf Taïk**[1,2]**, Golnoosh Farnadi**[1,2,3]

[1]Mila, Quebec AI Institute, Quebec, Canada,
[2]Université de Montréal, Quebec, Canada,
[3]McGill University, Quebec, Canada
nicola.neophytou@mila.quebec, afaf.taik@mila.quebec, farnadig@mila.quebec

Abstract

The aftermath of the Covid-19 pandemic saw more severe outcomes for racial minority groups and economically-deprived communities. Such disparities can be explained by several factors, including unequal access to healthcare, as well as the inability of low income groups to reduce their mobility due to work or social obligations. Moreover, senior citizens were found to be more susceptible to severe symptoms, largely due to age-related health reasons. Adapting vaccine distribution strategies to consider a range of demographics is therefore essential to address these disparities. In this study, we propose a novel approach that utilizes influence maximization (IM) on mobility networks to develop vaccination strategies which incorporate demographic fairness. By considering factors such as race, social status, age, and associated risk factors, we aim to optimize vaccine distribution to achieve various fairness definitions for one or more protected attributes at a time. Through extensive experiments conducted on Covid-19 spread in three major metropolitan areas across the United States, we demonstrate the effectiveness of our proposed approach in reducing disease transmission and promoting fairness in vaccination distribution.

Introduction

The fallout of Covid-19 revealed the stark inequalities in access to healthcare between social groups in diverse and urban areas (Ndugga et al. 2021; Price-Haywood et al. 2020; Millett et al. 2020; Joseph et al. 2020; Azar et al. 2020; Hsu et al. 2020; Gayam et al. 2021). Studies confirmed that economically-deprived communities and racial minorities experienced higher rates of infection, hospitalization and mortality as a result of Covid-19 (Kirby 2020; Tai et al. 2022; Alcendor 2020). The reasons for this disparity form a long chain of events, with unequal access to healthcare between socioeconomic groups, and therefore racial groups, at the root of it. Furthermore, studies on mobility networks in the US also revealed how minority communities were less able to reduce their mobility as quickly during the pandemic, and as a result suffered higher rates of infection (Chang et al. 2020). This can be largely attributed to underprivileged groups assuming the roles of frontline and critical infrastructure work, and also living and working in more crowded circumstances. Inequalities in access to the internet and ease

of travelling to test and vaccination sites are also factors contributing to this discrepancy (Yee, Bajaj, and Stanford 2022). These data demonstrate inequities in receiving resources throughout the pandemic, which also extends to vaccination (Perry et al. 2021; Kates, Tolbert, and Michaud 2021; Ndugga et al. 2022; Bayati et al. 2022). In Figure 1, we demonstrate how racial minorities and lower income communities in three US metropolitan areas were less able to reduce their mobility as quickly when the lockdown was introduced.

This disparity motivates the need for a fair vaccination strategy that differs from the current technique. In this work, we investigate a collection of alternative vaccination strategies that consider both mobility and fairness. We leverage an approach called influence maximization (IM), a network science technique designed to detect the most influential members of social networks, typically used in applications such as viral marketing campaigns (Chen, Wang, and Wang 2010). We adapt this principle to instead detect the neighborhoods or communities which exhibit the largest influence on a mobility network in terms of disease propagation. Such communities are likely to include essential workers who are less able to reduce their mobility during lockdowns (Rasnača 2020; Nguyen et al. 2020).

We adapt our IM approach to achieve fairness in vaccine allocation for racial groups as well as groups of different social statuses. Moreover, older individuals may be less mobile but more at risk of severe outcomes when exposed to the disease. It is essential not to overlook this trade-off; we therefore also design a strategy designed to protect communities based on higher risk and vulnerability. To summarize, our contributions are as follows:

1. A novel community-level influence maximization approach for identifying impactful neighborhoods, aiding targeted vaccination against disease transmission.

2. Extension of influence maximization to mitigate infection disparities among racial and income-level communities.

3. Introduction of a competitive method, merging influence maximization with prioritizing older communities to reduce overall infections.

4. Empirical validation on mobility networks from three major US metropolitan areas, utilizing real aggregated
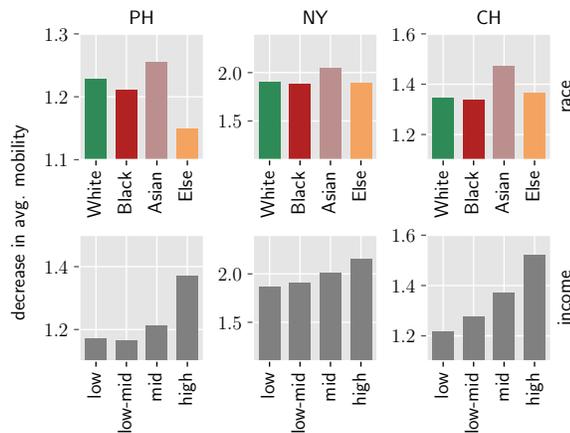
Figure 1: The *reduction* in mobility from before lockdown to during lockdown per racial groups and income groups in Philadelphia, New York and Chicago metropolitan areas. For all three areas, lower income groups and racial minorities belonging to lower income groups (see Fig. 2) were less able to reduce their mobility as quickly when transitioning to lockdown.

visit data from census block groups (CBGs) to points of interest (POIs) during the first five weeks of Covid-19 pandemic.

Related Work

Fairness in AI is an expanding area of research which has seen traction since the exposure of biases in several significant technologies (Mehrabi et al. 2021; Buolamwini and Gebru 2018; Angwin et al. 2016). This extends to the field of influence maximization, where various works impose fairness constraints on the optimization problem. For instance, Farnadi et al. provide a framework for applying a variety of fairness definitions to IM tasks (Farnadi, Babaki, and Gendreau 2020). Other works have included developing adversarial graph embeddings to achieve fair IM in social networks (Khajehnejad et al. 2020), and balancing majority and minority groups in IM when networks and diffusion processes demonstrate homophily (Anwar, Saveski, and Roy 2021). Ali et al. ensure fairness in exposure under a time-critical perspective, for example posting a job advertisement which should be reached by equal sub-populations before the deadline to apply for the position (Ali et al. 2021). This does not however consider the possibility of changing network structure over time, as is the case with mobility or contact networks used in epidemiology, and is therefore not extendable to the problem of vaccine distribution.

Meister et al. characterize communities by their social activity and vulnerability due to age in a comparative game for optimizing vaccination, however, they do not use real contact or mobility data to test their approach (Meister and Grycko 2021). Similarly, works which do consider demographic fairness in vaccination do not necessarily consider mobility to improve performance (Kipnis et al. 2021; National Academics of Sciences and Medicine 2020; Kadelka et al. 2022). Anahideh et al. propose a vaccine alloca-

tion solution which tackles the apparent trade-off between equal distribution amongst regions and demographic fairness (Anahideh, Kang, and Nezami 2022). Their approach defines effective distributions as those that cover many geographical areas; however, we argue that this does not necessarily reduce infections most effectively as it treats each area as equally influential in disease transmission. Commonly, works that do consider social contact often do not consider fairness implications (Jiangzuo et al. 2021). Much research considers how to optimize the age-based vaccination approach, but often other sensitive attributes are not considered (Wu, Wang, and Xu 2022; Sheldrick, Meyerowitz-Katz, and Tucker-Kellogg 2022; Kirwin et al. 2021; Ferranna, Cadarette, and Bloom 2021; Gozzi et al. 2022; Sheldrick, Meyerowitz-Katz, and Tucker-Kellogg 2022; Jentsch, Anand, and Bauch 2021; Buckner, Chowell, and Springborn 2021).

Similar to our approach, Minutoli et al. (Minutoli et al. 2020) lay the theoretical foundations for using IM to reduce disease transmission via vaccination, though they do not consider fairness. They use a simplistic propagation model which is not specific to a particular disease, nor does it consider important factors in disease transmission such as the number of people in a confined space, or the duration spent there, as we do. Additionally, their implementation is for contact networks containing interactions between individuals, rather than communities. We argue that privacy regulations on mobile tracking data will limit the ability to reconstruct such specific interactions, not to mention also matching demographic information to individuals in these interactions. Using aggregate visits as in our work is therefore more realistic in terms of the data available, and allows us to extract demographic information on the community level which we use for our fairness approaches. Rather than using large contact networks, condensed mobility networks of aggregate visits also drastically reduces the network size which is beneficial for computation.

Mehrab et al. use the same mobility data as in our contribution to guide vaccine distribution to groups with lower up-take rates (Mehrab et al. 2022). However, they use visit counts to public places to determine the best candidates, whereas we determine influence by simulating disease spread on top of the mobility network to predict infections. Further, their contribution is to identify groups with low vaccine uptake at a later stage of vaccine deployment, whereas we offer a solution for the first stages of the allocation, to ensure fairness from the beginning. To our knowledge, our work is the first approach to use *community-level* influence maximization to propose vaccination strategies which consider demographic fairness.

Preliminaries

Influence Maximization

IM is a network science technique used to identify the most influential nodes in a graph, with respect to their ability to propagate a certain transmissible quality, such as information or disease. IM is particularly popular in viral marketing problems on social media platforms, in which an entity

wants to share advertisements with only a few individuals online, but hopes to optimize this choice by selecting those who are more likely to share the information with the rest of the network.

While some techniques rely on heuristics like node centrality or betweenness to measure a node's importance in a graph (Newman 2005; Borgatti 2005), IM assumes access to a function which models the propagation of the given substance across the network. The algorithm uses this propagation function to generate its selection for the set of most influential nodes, often referred to as the seed set. In their seminal work, Kempe et al. (Kempe, Kleinberg, and Tardos 2015) proposed a greedy algorithm which achieves $(1 - \frac{1}{e})$ optimality guarantees on the seed set, so long as the objective function - which is the given propagation function - is both submodular and monotonic. However, the biggest drawback of the greedy strategy is its inefficient runtime. The commonly used propagation functions, such as the Linear Threshold (Chen, Yuan, and Zhang 2010) and Independent Cascade (Saito, Nakano, and Kimura 2008) models, are stochastic in nature. The greedy strategy, therefore, relies on running several Monte Carlo simulations of these stochastic propagation functions, which is costly. Much of the proceeding work following this focus on trying to improve runtime performance; CELF exploits the submodularity of the influence function to reduce the number of influence evaluations, achieving a runtime of up to 700 times faster than the greedy algorithm (Leskovec et al. 2007). In our experiments, we also use CELF to reduce the number of evaluations.

Mobility Networks

We draw on the work conducted by Chang et al. to construct our mobility networks and simulate Covid-19 propagation on them (Chang et al. 2020). The mobility network of a metropolitan statistical area (MSA) contains K nodes which represent CBGs, neighborhoods of a few hundred to a few thousand residents. Each CBG is denoted by c_i, with $i = 1, .., K$. The population of each CBG is known and is given by n_{c_i} for CBG c_i. The total number of residents in the network is, therefore, $N = \sum_{i=1}^{K} n_{c_i}$. Each individual in the network can belong to one of M social groups (e.g. racial groups) indexed by $j = 1, .., M$. In our work, we experiment with j representing racial groups and also groups based on their median household income, obtained from US Census data (Manson et al. 2022). For racial groups, each CBG can contain any number of individuals belonging to each social group, up to its total population size. The fraction of residents in CBG c_i who belong to racial group j is known and given by α_{ij}, where $0 \leq \alpha_{ij} \leq 1$, and therefore $\sum_{j=1}^{M} \alpha_{ij} = 1 \forall i$. The total number of individuals in the network belonging to racial group j is therefore obtained by $N_j = \sum_{i=1}^{K} \alpha_{ij} n_{c_i}$. For social status groups, we split the CBGs into M groups defined by the median income of the households in that CBG. The total number of individuals belonging to a group is, therefore, $N_j = \sum_{i=1}^{K} n_{c_i}$ if c_i belongs to income group j.

Covid-19 Propagation Model

We use the Covid-19 simulation proposed by Chang et al. (Chang et al. 2020), which models the propagation of the disease on the network of CBGs and POIs. The constructed hourly visits from the network describe how many individuals travel from CBGs to POIs per hour. POIs here represent public places such as restaurants, gyms and religious centers, where interactions with members of other communities can happen, and the disease can spread accordingly. Individuals can also transmit the disease amongst members of their home CBG. The number of new infections per CBG is therefore a summation of two terms drawn from different distributions; one Poisson distribution for new exposures from POIs, and one Binomial distribution for new exposures from their home CBG and other places which may not be accounted for in the POIs, such as public transportation. The previous work fixed the parameters of the model by comparing it to the counts of real Covid-19 cases, which we also reproduce in our work. Our vaccination approach can technically be used to combat any disease, simply by swapping the influence function σ for a model of said disease. This would mean changing the parameters of the model when calibrating it to real case counts. Further details of the model and its calibration can be found in the full version of our paper (Neophytou, Taïk, and Farnadi 2024).

At all time steps in the disease simulation, we have access to the number of susceptible, infected, exposed, and recovered or removed individuals residing in each CBG. We maintain a vector of size K for each of these states, denoted respectively by S, E, I, and R. Their elements are indexed by i and contain the fraction of individuals in CBG c_i belonging to that state. For example, element I_i^t contains the fraction of individuals in CBG c_i who are in the infected state at time t. In this work, we are only interested in the final rates of exposed, infected or recovered/removed (EIR) individuals in the final time step of the model, T, and readers can assume we use the final iteration at T of these vectors from here onwards. For example, the sum of all exposed-or-worse individuals in the network at time T is given by $N_{EIR} = \sum_{i=1}^{K} (E_i^T + I_i^T + R_i^T) n_{c_i}$.

Individuals in the network therefore always belong to one of M social groups, as well as one of the four SEIR states. Since we only have access to these statistics on the community level, we approximate the number of people in the network belonging to racial group j and exposed-or-worse as $N_{EIR_j} = \sum_{i=1}^{K} (E_i^T + I_i^T + R_i^T) \alpha_{ij} n_{c_i}$. For income group j, the equivalent is obtained by $N_{EIR_j} = \sum_{i=1}^{K} (E_i^T + I_i^T + R_i^T) n_{c_i}$ for CBGs belonging to income group j.

Proposed Approach

In this section, we present three methods of targeted vaccination using IM. Firstly, we present our simple method for vaccinating with IM and no additional constraints. We then outline two methods for introducing fairness to IM, for both racial groups and income groups. Finally, we present a method for applying weights to communities corresponding to their relative risk, in order to use IM and still prioritize

older communities that are more vulnerable to severe outcomes if infected.

In our approach, the treatment of the three sensitive attributes (also referred to as protected attributes) in the network - race, income level and age - are not the same. For the racial and income groups, we aim to achieve fairness according to their population size in the network. An individual should be no more at risk of infection due to their race or income than what is expected given the racial or income group's population size in the network. However, the same strategy should not be adopted for age, since infection can lead to more severe outcomes for older individuals, making them more high risk. Therefore, we strive to achieve fairness amongst race and income, but adopt a *bias* with respect to age, in order to protect individuals at high-risk.

As addressed in the Preliminaries, the greedy approach in influence maximization provides provable guarantees on the optimality of the seed set, so long as the influence function is both monotonic and submodular. Previous findings have found issues proving these properties in a temporal SIR model (Erkol, Mazzilli, and Radicchi 2022). Similar to this work, we argue that the greedy approach is still effective despite the propagation model exhibiting submodularity violations. Additionally, the greedy approach is more scalable for greater network sizes than solving the optimization problem exactly, particularly when fairness constraints are required (Farnadi, Babaki, and Gendreau 2020). Further, we argue that the greedy approach provides an element of model interpretability, which is particularly important when justifying why one neighborhood should receive vaccines over another; in our case, we can clearly demonstrate with the greedy approach that a neighborhood with higher mobility and influence is more likely to be selected for vaccination. Further, we argue that using community-level influence maximization, with aggregated data on CBGs rather than individuals, is a more realistic approach when we want to obtain demographic information for fairness purposes. Privacy concerns (rightly) limit access to fine-grained data on individual mobility and their sensitive attributes, but here we use Census data to match demographic data to CBGs.

Vaccinating With Influence Maximization

In this section, we outline how to select the most influential communities in the network in terms of disease spread using IM and CELF. The algorithm is provided in Algorithm 1. We begin with a budget B corresponding to the number of vaccines available for allocation to the whole network. We can simulate the disease spread from a set Z of CBGs, in order to quantify how influential those communities are. In the general case, Z can contain any number of CBGs, but to quantify the influence of just one, only that CBG would be contained in the set Z, e.g. $Z = \{c_i\}$ for $i = 1, .., K$. For all K CBGs in the network, we then conduct simulations of the disease spreading from that neighborhood alone. We maintain a list of lists $L = [[1, \sigma(\{c_1\})], ..., [K, \sigma(\{c_K\})]]$, where for CBG i, $\sigma(\{c_i\})$ represents the N_{EIR} count as a result of simulating disease spread starting only from c_i. L, therefore, contains the list of pairs of the candidate CBGs indices along with their corresponding influence, as a count

---

Algorithm 1: Selecting CBGs to vaccinate using IM and CELF

Input: budget B, number of CBGs K, disease model σ
Output: CBGs to vaccinate V

1: $B' \leftarrow 0, spread \leftarrow 0, L \leftarrow [], Z \leftarrow []$
2: **for** $i = 1, \ldots, K$ **do**
3:    $L.\text{append}([i, (\sigma(\{c_i\}) - spread)/n_{c_i}])$
4: **end for**
5: sort L by gain, descending
6: $Z.\text{append}(L_{0,0})$ {add cbg with best gain to Z}
7: $spread \leftarrow L_{0,1} * n_{Z_{-1}}$
8: $B' \leftarrow n_{Z_{-1}}$ {update budget used}
9: **while** there are possible candidates in L **do**
10:    $matched \leftarrow False$
11:    **while** not matched **do**
12:       $best \leftarrow L_{0,0}$
13:       $L_{0,1} \leftarrow (\sigma(Z \cup \{best\}) - spread)/n_{best}$
14:       sort L by gain, descending
15:       $matched \leftarrow L_{0,0}$ is $best$
16:    **end while**
17:    $spread \leftarrow spread + L_{0,1} * n_{best}$
18:    $Z.\text{append}(L_{0,0})$
19:    $B' \leftarrow B' + n_{Z_{-1}}$ {update budget used}
20:    $L \leftarrow L[1 :]$ {remove best from the candidate list}
21:    keep only candidates in L which cannot exceed B
22: **end while**
23: **return** Z

---

of how many people resulted in exposed-or-worse states. We initialize the set of nodes to vaccinate, Z, as an empty set. Then, in each iteration, the CBGs with the greatest marginal gain are greedily added to Z. The marginal gain is the difference in influence between the current set of selected CBGs ($spread$), and the influence of the current selected CBGs *plus a potential candidate CBG*. Note that the gain is also normalized by the population of the CBG, n_{c_i}. We implement this normalized version of IM since we want to select CBGs that are the most influential per their population, and CBGs with a higher population use more of the vaccination budget than CBGs with a lower population. We keep track of how much budget is used so far with B', which gets updated with the population sizes of CBGs when they are added to Z.

In lines 9-15, we perform a check to test whether the highest-influence candidate after the previous iteration is still the highest-influence candidate in the current iteration. If this is true, we omit the requirement to re-calculate the influence of the other candidates. This exploits the submodularity of the influence function, since the marginal gain of adding CBG c_i to a smaller set Z can only decrease. This is the contribution made by CELF (Leskovec et al. 2007), which we use to improve run time. The algorithm then continues to add candidate CBGs to Z so long as their addition does not exceed the budget B. The final output of the model is therefore a set of CBGs to be vaccinated, which we call V. Our subsequent variations of this contribution in the next sections adapt this method to apply demographic fairness.

IM With Equal Treatment

Equal treatment is an existing fairness notion in the domain of fair IM, which aims to achieve fair representations of social groups in the final set of selected nodes V. This is equivalent to achieving the same demographic distribution in the set of communities to be selected for vaccination as in the whole network.

We model this task as a multiple knapsack problem (Chekuri and Khanna 2005), whereby each social group $j = 1, .., M$ is allocated a number of vaccines based on the fraction of their population in the network. Each social group j, therefore, receives its own budget B_j, corresponding to the number of vaccines to be allocated to the group, given by

$$B_j = \frac{N_j}{N} B \quad (1)$$

where N_j is the number of individuals in the network belonging to group j and N is the total network population. To implement this, when a CBG is selected for vaccination, we update the budget used by each of the M social groups, between lines 19 and 20 of Algorithm 1. Additionally, after line 21, we perform another check to ensure the remaining candidate list contains only CBGs whose addition would not violate any of the social group budgets B_j. We use these definitions to outline two strategies: equal treatment by racial groups and equal treatment by median household income.

Equal Treatment by Racial Groups (IM-R) When performing equal treatment for racial groups, the budget of racial group j is updated when CBG c_i is selected for vaccination via

$$B'_j = B'_j + n_{c_i} \alpha_{ij} \quad (2)$$

This update is performed for all M racial groups when any additional CBG is selected.

We acknowledge that our fairness framework takes a Western-centric perspective, particularly with respect to race, and is relevant mostly to countries containing urban areas with high diversity, as is more typical of the global West. The most effective strategy for each country, however, will not be the same (Eyal et al. 2022). We therefore propose IM-I to counteract this bias; many non-Western countries may not have the same extent of racial diversity, but will still experience income disparities in their urban areas. Using income as a sensitive attribute therefore still provides a fair IM method which is relevant for areas with low racial diversity. Further, Figure 2 identifies that in our selected MSAs, the White population tends to dominate higher income CBGs, while historically marginalized groups of Black or African-American are more prevalent in lower income groups, as is typical of high-diversity urban areas in the West. Therefore, fairness by income level may also achieve fairness by racial groups for our selected MSAs.

Equal Treatment by Median Income (IM-I) We perform a similar equal treatment scenario, this time with social groups defined by income. We use labels of the median household income of each CBG. The distribution of the CBG median income is split into four quartiles. We then bucket the CBGs into one of $M = 4$ groups according to
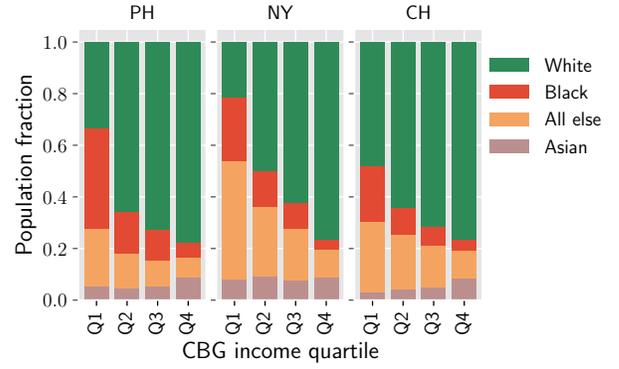

Figure 2: Racial distributions of CBGs grouped by their median income. Income groups are determined by quartiles of the median income distribution. Results are for three MSAs: Philadelphia, New York and Chicago.

which quartile its median income falls into. The budget is split in the same way as with race, using $B_j = \frac{N_j}{N} B$ for each income group j, where N_j is the total population of that income group. However this time, the budget updates are given by

$$Bj' = \begin{cases} B'_j + n_{c_i}, & \text{if } c_i \text{ belongs to group } j \\ B'_j, & \text{otherwise} \end{cases} \quad (3)$$

IM With Age-Associated Risk-Weights (IM-A)

While older individuals in the network are less mobile, they are more likely to experience severe consequences if they contract the disease, including hospitalisation and death. As a result, it is important to consider this tradeoff when using IM. To do so, we incorporate this notion into our IM technique such that the significance of infecting a person from a CBG with a higher median age is greater. We implement this by weighting the CBGs with a "risk-factor" according to their median age. The purpose is to not only select CBGs for vaccination that are highly influential in the network but in particular, find those CBGs that pose more of a risk of exposing older communities to the disease. We achieve this by scaling the influence calculations σ, which is used as the selection criteria for vaccination (see Algorithm 1). We construct a vector μ of size K containing the associated risk-weights for each CBG, determined by its median age. We fill this vector with the death rates per age group from the CDC, which are the rates of death of older groups compared to 18-29 year olds in the US (CDC 2023). We report these values in our full paper (Neophytou, Taïk, and Farnadi 2024). The resulting influence, which before was just the sum of exposed-or-worse individuals in the network, is now instead a weighted sum where each CBG is weighted according to its risk-factor in vector μ, given by

$$\sigma_A = \sum_{i=1}^{K} \mu_i n_{c_i}(E_i + I_i + R_i). \quad (4)$$

We use this metric as a proxy to account for infections as well as more severe cases and deaths. In Algorithm 1, the influence function σ is replaced by σ_A in lines 3 and 13.

Multiple Protected Attributes

The methods proposed so far each strives for fairness using one protected attribute at a time. However, it is possible that achieving the desired fairness for one protected attribute is not beneficial for another. For example, older communities tend to be predominantly White in the US. Therefore, a vaccination strategy based solely on age, without consideration for fair distribution amongst racial groups, leads to an unfair allocation. This would also lead to unfair allocation amongst income groups as White populations tend to dominate higher income CBGs, as shown in Figure 2. To address this issue, we propose the following combinations of our methods:

IM With Race Groups and Age-Associated Risk-Weights (IM-RA) We perform equal treatment to achieve representative allocation of vaccines amongst racial groups, but use the influence function σ_A to apply a heavy penalty for infecting older communities.

IM With Median Income and Age-Associated Risk-Weights (IM-IA) Similarly, we perform equal treatment of vaccines amongst the four income groups and replace the influence function with σ_A. In our experiments, we later find high numbers of submodularity violations when using the σ_A influence function. For all experiments using the "-A" suffix, we, therefore, omit the recalculation checks in Lines 10 to 15 of Algorithm 1.

Experiments

Dataset We conduct experiments on three mobility networks of MSAs in the US, constructed from individual mobile tracking data from SafeGraph, from the Dewey platform (SafeGraph 2023). We use the implementation proposed by (Chang et al. 2020) to construct these networks, and also use their Covid-19 model as our influence function. A mobility network is constructed as a temporal bipartite graph $G^t = (V, E^t)$, whose nodes V are either CBGs, which are communities of between 600 and 3,000 US residents, or POIs. A directed weighted edge $e^t(v_1, v_2) \in E^t$ represents the number of residents from a CBG who are visiting a POI at hour t. The graph varies over time, such that no nodes are added or removed, but the edge weights vary hourly. We use five weeks of visits from CBGs to POIs beginning on the 2nd of March 2020. For our implementation, only the first two weeks are used to select the most influential communities, and vaccination is implemented after the two-week mark. We used data from the beginning of March in order to exploit the full mobility of individuals before lockdown. Otherwise, it would not have been possible to measure the effect of the proposed vaccination strategies separately from the effect of lockdown.

Throughout the experiments, we set the vaccination budget to 5% of the population size of the whole network. We experiment with three MSAs - Philadelphia, New York and Chicago - each of which encompasses the main city as well as the wider metropolitan area. These particular MSAs were selected based on their high racial diversity, and high discrepancy in infections between racial groups as reported in the previous work. The outputs are mobility networks

modelling metropolitan areas of populations between 6 and 10 million residents. We provide details of further pre-processing steps and network statistics in the full paper.

Baselines Each vaccination strategy is run over 30 random seeds, and we report the average results in comparison to three baseline approaches. We design our own baselines here, since the closest approach to ours which uses IM for vaccination uses a generic propagation function which is not specific to a particular disease, and is valid only on rooted trees and not general graphs (Minutoli et al. 2020).

- *No vaccination:* We model free-spreading Covid-19 during the total five-week period, without any vaccination strategy.
- *Random vaccination (RAND):* We implement a random selection of CBGs for vaccination within the budget B. We collect results over three random seeds and report the average.
- *Current strategy proxy (CS):* Here we replicate the current strategy which prioritizes older communities. We select the oldest communities for vaccination, by the median age of the CBG, up to the vaccine budget B.

Fairness Evaluation Metrics In addition to evaluating the performance of the vaccination strategies, we propose two methods of evaluating fairness for social groups in this context. In both cases, we measure the discrete KL-divergence between two distributions; we compare a distribution from the outcome of our experiments $p(j)$ to an ideal "fair" distribution $q(j)$. We draw on fairness notions from the IM literature - equal treatment and equal outcomes. For both measures, the fair distribution $q(j)$ corresponds to the fractions of each social group in the network, $q(j) = N_j/N$.

- **Equal Treatment** For equal treatment, we aim to obtain a fair representation of each social group j within the CBGs selected for vaccination. As such, the output distribution $p(j)$ is the proportion of social group j amongst vaccinated CBGs, $p(j) = (N_j/N)_V$.
- **Equal Outcome** To obtain equal outcomes, the goal is to ensure that no individual is more at risk of infection than the fraction of that social group in the network dictates. In this case, we set $p(j)$ to be the proportion of infections received by social group j, as a result of our vaccination strategy. This can be written as $p(j) = (N_j/N)_{EIR}$.

Results and Discussion

Reducing Overall Infections The performance of each vaccination strategy in reducing the number of infections is shown in Figure 3 (green). For every MSA, all vaccination strategies using IM outperform both RAND and CS. Though the infections decrease by only a few percent, given the network size, these percentage point differences are significant. For example, a 5% decrease in infections for Philadelphia corresponds to around 18,000 fewer people infected (including estimates of unreported infections). We observe that the variations of IM experiments which include fairness (the last five bars) do not experience a significant decrease in performance even when optimizing for both performance *and* fair-
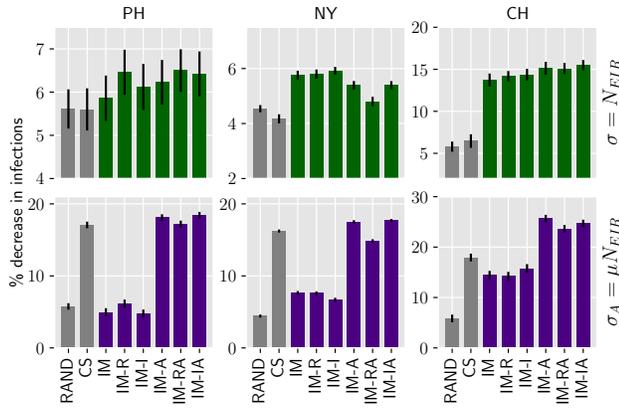
Figure 3: Performance measured by percentage decrease in infections (top), and percentage decrease in risk-weighted infections, i.e. with a weighted penalty of infecting older communities (bottom), compared to not vaccinating. Higher is better for both metrics.
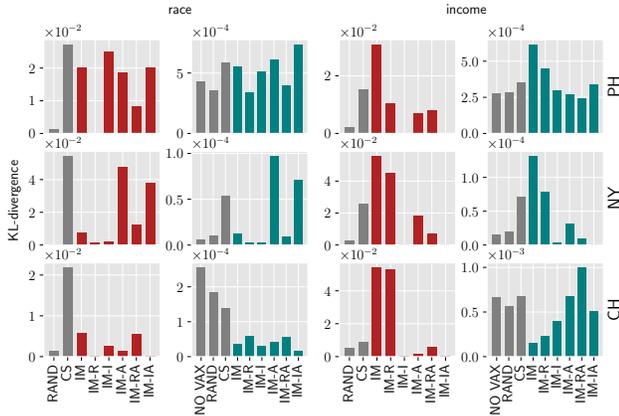

Figure 4: The KL-divergence scores measure fair treatment (red) and fair outcomes (blue) with respect to racial groups (left) and income groups (right). Lower D_{KL} corresponds to better fairness for both metrics.

ness. This illustrates that fairness considerations in vaccination distribution do not have to come at the cost of increasing infection counts.

Infections in High-Risk Groups   Figure 3 (purple) reports the percentage decrease in infections weighted by age-associated risk, as described in Equation 4. The experiments optimizing for age are the baseline CS and our contributions IM-A, IM-RA, and IM-IA, which therefore perform best for this metric. However, we see that all of our proposed solutions which optimize for age outperform the current strategy, even when they are also optimizing for another sensitive attribute at the same time. The results testify to several alternative solutions that better protect older communities as well as ensure fairness for other sensitive attributes like race and income.

Comparing Fairness Notions   In Figure 4 left, we present results for equal treatment (red) and equal outcome (blue) for racial groups, and in Figure 4 right we present the

same metrics for income groups. Though the D_{KL} values are small, these still correspond to significant differences in these large networks. For example, the current strategy model (CS) has a D_{KL} of around 5.8×10^{-4} for Philadelphia, which corresponds to the Black population suffering around 6,000 more infections than if they were distributed according to their proportion in the network. Critically, we see that IM-R methods successfully achieve equal treatment for racial groups across all MSAs, as does IM-RA, with the exception of Chicago. The effect is stronger for income, where IM-I and IM-IA experiments achieve near-perfect (D_{KL} of zero) distribution of vaccines to income groups, for all MSAs. Though equal treatment is important to ensure, we are more interested in an equal outcome, as infections are the more serious consequence. We see that the experiments which perform best for equal outcomes amongst races are also IM-R and IM-RA. This demonstrates that achieving equal treatment can be very effective in delivering fair outcomes for those same demographic groups. Additionally, for some MSAs, constraining on fairness by *income* (IM-I) leads to fair outcomes for *race*, and vice versa. This implies that the objectives of achieving fairness amongst races and fairness amongst social status are similar.

Optimizing for Multiple Sensitive Attributes   Since the results of every vaccination experiment for each MSA can differ, it is possible that there is no one one-size-fits-all best vaccination strategy for every urban area. In particular, we can identify how accounting for higher-risk individuals (experiments with the "-A" suffix) can work favourably for achieving demographic fairness in infections for some but not all MSAs (see Chicago, Figure 4 blue, left and New York, Figure 4 blue, right). Despite this, we can identify at least one strategy per metropolitan area which achieves high performance in reducing infections overall, as well as a competitive result for the infection outcomes of *all three* sensitive attributes of age, race, and social status by income: IM-RA for Philadelphia and New York, and IM-IA for Chicago.

Conclusion

For policy-makers, choosing a vaccination strategy amongst those presented here is non-trivial. There is no one-size-fits-all solution for every urban area. However, we demonstrate that, for all networks we tested here, one of our proposed methods can successfully ensure demographic fairness for all three sensitive attributes. We, therefore, argue that community-level influence maximization should be incorporated into whichever ethical stance is taken, and we present the methodology to do so.

Our approach can be extended to accommodate multiple rounds of vaccine allocation, as commonly observed in real-world scenarios. It would be necessary to capture the mobility shifts at different stages of lockdown, and how this affects demographic groups differently. There are many factors which could be incorporated to make the simulation more realistic, such as vaccine hesitancy. For later stages of vaccine roll-out, our approach could be combined with data on hesitancy rates to prioritize neighborhoods with low up-take, while maintaining demographic fairness.

Acknowledgements

References

Alcendor, D. J. 2020. Racial Disparities-Associated COVID-19 Mortality among Minority Populations in the US. *Journal of Clinical Medicine*, 9.

Ali, J.; Babaei, M.; Chakraborty, A.; Mirzasoleiman, B.; Gummadi, K.; and Singla, A. 2021. On the Fairness of Time-Critical Influence Maximization in Social Networks. *IEEE Transactions on Knowledge and Data Engineering*.

Anahideh, H.; Kang, L.; and Nezami, N. 2022. Fair and diverse allocation of scarce resources. *Socio-Economic Planning Sciences*, 80.

Angwin, J.; Larson, J.; Mattu, S.; and Kirchner, L. 2016. Machine Bias. *ProPublica*.

Anwar, M. S.; Saveski, M.; and Roy, D. 2021. Balanced Influence Maximization in the Presence of Homophily. *WSDM 2021 - Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 175–183.

Azar, K. M.; Shen, Z.; Romanelli, R. J.; Lockhart, S. H.; Smits, K.; Robinson, S.; Brown, S.; and Pressman, A. R. 2020. Disparities in outcomes among COVID-19 patients in a large health care system in California. *Health Affairs*, 39(7): 1253–1262.

Bayati, M.; Noroozi, R.; Ghanbari-Jahromi, M.; and Jalali, F. S. 2022. Inequality in the distribution of Covid-19 vaccine: a systematic review. *International Journal for Equity in Health*.

Borgatti, S. P. 2005. Centrality and network flow. *Social Networks*, 27: 55–71.

Buckner, J. H.; Chowell, G.; and Springborn, M. R. 2021. Dynamic prioritization of COVID-19 vaccines when social distancing is limited for essential workers. *Proceedings of the National Academy of Sciences*.

Buolamwini, J.; and Gebru, T. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR*.

CDC. 2023. Risk for COVID-19 Infection, Hospitalization, and Death By Age Group. https://archive.cdc.gov/#/details?url=https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-age.html. Last Accessed: 2023-01-24.

Chang, S.; Pierson, E.; Koh, P. W.; Gerardin, J.; Redbird, B.; Grusky, D.; and Leskovec, J. 2020. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature 2020 589:7840*, 589(7840): 82–87.

Chekuri, C.; and Khanna, S. 2005. A Polynomial Time Approximation Scheme for the Multiple Knapsack Problem. *SIAM Journal on Computing*, 35.

Chen, W.; Wang, C.; and Wang, Y. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD '10: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*.

Chen, W.; Yuan, Y.; and Zhang, L. 2010. Scalable Influence Maximization in Social Networks under the Linear Threshold Model. *IEEE International Conference on Data Mining*.

Erkol, S.; Mazzilli, D.; and Radicchi, F. 2022. Effective submodularity of influence maximization on temporal networks. *Physical Review E*, 106.

Eyal, N.; Gheaus, A.; Gosseries, A.; Magalhaes, M.; Ngosso, T.; Steuwer, B.; Tangcharoensathien, V.; Trifan, I.; and Williams, A. 2022. Coronavirus Disease 2019 (COVID-19) Vaccine Prioritization in Low- and Middle-Income Countries May Justifiably Depart From High-Income Countries' Age Priorities. *Clinical Infectious Diseases*.

Farnadi, G.; Babaki, B.; and Gendreau, M. 2020. A Unifying Framework for Fairness-Aware Influence Maximization. *WWW '20: Companion Proceedings of the Web Conference 2020*.

Ferranna, M.; Cadarette, D.; and Bloom, D. E. 2021. COVID-19 Vaccine Allocation: Modeling Health Outcomes and Equity Implications of Alternative Strategies. *Engineering (Beijing, China)*.

Gayam, V.; Chobufo, M. D.; Merghani, M. A.; Lamichhane, S.; Garlapati, P. R.; and Adler, M. K. 2021. Clinical characteristics and predictors of mortality in African-Americans with COVID-19 from an inner-city community teaching hospital in New York. *Journal of medical virology*, 93(2): 812–819.

Gozzi, N.; Chinazzi, M.; Dean, N. E.; Longini, I. M.; Halloran, E.; Perra, N.; and Vespignani, A. 2022. Estimating the impact of COVID-19 vaccine allocation inequities: a modeling study. *medRxiv*.

Hsu, H. E.; Ashe, E. M.; Silverstein, M.; Hofman, M.; Lange, S. J.; Razzaghi, H.; Mishuris, R. G.; Davidoff, R.; Parker, E. M.; Penman-Aguilar, A.; Clarke, K. E.; Goldman, A.; James, T. L.; Jacobson, K.; Lasser, K. E.; Xuan, Z.; Peacock, G.; Dowling, N. F.; and Goodman, A. B. 2020. Race/Ethnicity, Underlying Medical Conditions, Homelessness, and Hospitalization Status of Adult Patients with COVID-19 at an Urban Safety-Net Medical Center - Boston, Massachusetts, 2020. *MMWR. Morbidity and mortality weekly report*, 69(27): 864–869.

Jentsch, P. C.; Anand, M.; and Bauch, C. T. 2021. Prioritising COVID-19 vaccination in changing social and epidemiological landscapes: a mathematical modelling study. *The Lancet Infectious Diseases*.

Jiangzuo, C.; Hoops, S.; Marathe, A.; Mortveit, H.; Lewis, B.; Venkatramanan, S.; Haddadan, A.; Bhattacharya, P.; Adiga, A.; Vullikanti, A.; Srinivasan, A.; Wilson, M.; Ehrlich, G.; Fenster, M.; Eubank, S.; Barrett, C.; and Marathe, M. 2021. Prioritizing allocation of COVID-19 vaccines based on social contacts increases vaccination effectiveness. Preprint available on medRxiv.

Joseph, N. P.; Reid, N. J.; Som, A.; Li, M. D.; Hyle, E. P.; Dugdale, C. M.; Lang, M.; Betancourt, J. R.; Deng, F.; Mendoza, D. P.; Little, B. P.; Narayan, A. K.; and Flores, E. J. 2020. Racial and Ethnic Disparities in Disease Severity on Admission Chest Radiographs among Patients Admitted with Confirmed Coronavirus Disease 2019: A Retrospective Cohort Study. *Radiology*, 297(3): E303–E312.

Kadelka, C.; Islam, M. R.; McCombs, A.; Alston, J.; and Morton, N. 2022. Ethnic homophily affects vaccine prioritization strategies. *Journal of Theoretical Biology*.

Kates, J.; Tolbert, J.; and Michaud, J. 2021. The COVID-19 "Vaccination Line": An Update on State Prioritization Plans.

Kempe, D.; Kleinberg, J.; and Tardos, É. 2015. Maximizing the spread of influence through a social network. *Theory of Computing*, 11.

Khajehnejad, M.; Asgharian Rezaei, A.; Babaei, M.; Hoffmann, J.; Jalili, M.; and Weller, A. 2020. Adversarial Graph Embeddings for Fair Influence Maximization over Social Networks. In *IJCAI'20: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*.

Kipnis, P.; Soltesz, L.; Escobar, G. J.; Myers, L.; and Liu, V. X. 2021. Evaluation of Vaccination Strategies to Compare Efficient and Equitable Vaccine Allocation by Race and Ethnicity Across Time. *JAMA Health Forum*.

Kirby, T. 2020. Evidence mounts on the disproportionate effect of COVID-19 on ethnic minorities. *The Lancet Respiratory Medicine*, 8.

Kirwin, E.; Rafferty, E.; Harback, K.; Round, J.; and McCabe, C. 2021. A Net Benefit Approach for the Optimal Allocation of a COVID-19 Vaccine. *PharmacoEconomics*.

Leskovec, J.; Krause, A.; Guestrin, C.; Faloutsos, C.; VanBriesen, J.; and Glance, N. 2007. Cost-effective outbreak detection in networks. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Manson, S.; Schroeder, J.; Van Riper, D.; Kugler, T.; and Ruggles, S. 2022. IPUMS National Historical Geographic Information System: Version 17.0 [dataset] Minneapolis, MN: IPUMS. http://doi.org/10.18128/D050.V17.0.

Mehrab, Z.; Wilson, M. L.; Chang, S.; Harrison, G.; Lewis, B.; Telionis, A.; Crow, J.; Kim, D.; Spillmann, S.; Peters, K.; Leskovec, J.; and Marathe, M. V. 2022. Data-Driven Real-Time Strategic Placement of Mobile Vaccine Distribution Sites. *The Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22)*.

Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; and Galstyan, A. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54: 1–35.

Meister, H.; and Grycko, E. 2021. Fair Distribution of Vaccines Among Population Groups. Https://doi.org/10.18445/20210623-151738-0.

Millett, G. A.; Jones, A. T.; Benkeser, D.; Baral, S.; Mercer, L.; Beyrer, C.; Honermann, B.; Lankiewicz, E.; Mena, L.; Crowley, J. S.; Sherwood, J.; and Sullivan, P. S. 2020. Assessing differential impacts of COVID-19 on black communities. *Annals of epidemiology*, 47: 37–44.

Minutoli, M.; Prathyush, S.; Halappanavar, M.; Tumeo, A.; Kalyananaraman, A.; and Vullikanti, A. 2020. PREEMPT: Scalable Epidemic Interventions Using Submodular Optimization on Multi-GPU Systems. In *SC20: International Conference for High Performance Computing, Networking, Storage and Analysis*.

National Academics of Sciences, E.; and Medicine. 2020. A Framework for Equitable Allocation of COVID-19 Vaccine. In *Framework for equitable allocation of Covid-19 vaccine*. Washington, DC: The National Academies Press.

Ndugga, N.; Hill, L.; Artiga, S.; and Haldar, S. 2022. Latest Data on COVID-19 Vaccinations by Race/Ethnicity — KFF.

Ndugga, N.; Pham, O.; Hill, L.; Samantha, A.; and Mengistu, S. 2021. Early State Vaccination Data Raise Warning Flags for Racial Equity — KFF.

Neophytou, N.; Taïk, A.; and Farnadi, G. 2024. Promoting Fair Vaccination Strategies through Influence Maximization: A Case Study on COVID-19 Spread. *arXiv*.

Newman, M. J. 2005. A measure of betweenness centrality based on random walks. *Social Networks*, 39–54.

Nguyen, L. H.; Drew, D. A.; Graham, M. S.; Joshi, A. D.; Guo, C.-G.; et al. 2020. Risk of COVID-19 among front-line health-care workers and the general community: a prospective cohort study. *The Lancet Public Health*, 5(9): e475–e483.

Perry, M.; Akbari, A.; Cottrell, S.; Gravenor, M. B.; Roberts, R.; Lyons, R. A.; Bedston, S.; Torabi, F.; and Griffiths, L. 2021. Inequalities in coverage of COVID-19 vaccination: A population register based cross-sectional study in Wales, UK. *Vaccine*, 39(42): 6256–6261.

Price-Haywood, E. G.; Burton, J.; Fort, D.; and Seoane, L. 2020. Hospitalization and Mortality among Black Patients and White Patients with Covid-19. *The New England journal of medicine*, 382(26): 2534–2543.

Rasnača, Z. 2020. Essential but unprotected: highly mobile workers in the EU during the COVID-19 pandemic. *ETUI Research Paper-Policy Brief*, 9.

SafeGraph. 2023. SafeGraph Data. https://www.safegraph.com. Last Accessed: 2023-12-18.

Saito, K.; Nakano, R.; and Kimura, M. 2008. Prediction of Information Diffusion Probabilities for Independent Cascade Model. *Knowledge-Based Intelligent Information and Engineering Systems*.

Sheldrick, K. A.; Meyerowitz-Katz, G.; and Tucker-Kellogg, G. 2022. Plausibility of Claimed Covid-19 Vaccine Efficacies by Age: A Simulation Study. *American Journal of Therapeutics*.

Tai, D. B. G.; Sia, I. G.; Doubeni, C. A.; and Wieland, M. L. 2022. Disproportionate Impact of COVID-19 on Racial and Ethnic Minority Groups in the United States: a 2021 Update. *Journal of Racial and Ethnic Disparities*.

Wu, H.; Wang, K.; and Xu, L. 2022. How can age-based vaccine allocation strategies be optimized? A multi-objective optimization framework. *Frontiers in Public Health*.

Yee, V.; Bajaj, S. S.; and Stanford, F. C. 2022. Building a pandemic supply chain — equity over equality. *Nature Medicine 2022 28:4*, 28(4): 609–610.