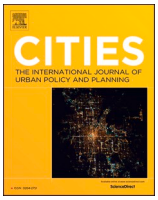
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Cities 128 (2022) 103805

Contents lists available at ScienceDirect

Cities

journal homepage: www.elsevier.com/locate/cities



Data-driven contact network models of COVID-19 reveal trade-offs between costs and infections for optimal local containment policies

Chao Fan [a,*], Xiangqi Jiang [b], Ronald Lee [b], Ali Mostafavi [a,*]

[a] Zachry Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77843-3136, United States of America
[b] Department of Computer Science and Engineering, Texas A&M University, College Station, TX 77843-3112, United States of America

ARTICLE INFO

Keywords:
Urban policy
Local containment
Pandemic
Data-driven network models

ABSTRACT

While several non-pharmacological measures have been implemented for a few months in an effort to slow the coronavirus disease (COVID-19) pandemic in the United States, the disease remains a danger in a number of counties as restrictions are lifted to revive the economy. Making a trade-off between economic recovery and infection control is a major challenge confronting many hard-hit counties. Understanding the transmission process and quantifying the costs of local policies are essential to the task of tackling this challenge. Here, we investigate the dynamic contact patterns of the populations from anonymized, geo-localized mobility data and census and demographic data to create data-driven, agent-based contact networks. We then simulate the epidemic spread with a time-varying contagion model in ten large metropolitan counties in the United States and evaluate a combination of mobility reduction, mask use, and reopening policies. We find that our model captures the spatial-temporal and heterogeneous case trajectory within various counties based on dynamic population behaviors. Our results show that a decision-making tool that considers both economic cost and infection outcomes of policies can be informative in making decisions of local containment strategies for optimal balancing of economic slowdown and virus spread.

1. Introduction

The coronavirus disease 2019 (COVID-19) has caused a global pandemic threatening public health and human well-being (Headey et al., 2020). As of November 4, 2020, 47.7 million confirmed cases and 1.2 million deaths have been reported worldwide (John Hopkins University, 2020). In response to the pandemic and in the absence of effective vaccines and drugs, numerous non-pharmaceutical interventions, such as stay-at-home and social distancing orders, have been implemented in the majority of the affected countries and regions (Chinazzi, 2020; Kraemer, 2020). For example, Harris County in Texas enacted stay-at-home orders starting March 2020, and mandatory mask order in May 2020. These measures were enacted to quarantine infected persons thus isolating cases and suppressing virus transmission (Rader et al., 2020). The execution of these measures, however, caused a significant economic slowdown (Polyakova et al., 2020). In particular, travel restrictions and business lockdowns have caused soaring unemployment and a decline in tax revenue (Auray & Eyquem, 2020). Hence, it is essential to examine policy solutions for making an optimal trade-off between virus control and economic recovery, especially in hard-hit regions.

As society faces this global public health challenge, numerous studies in a variety of research streams relevant to pandemic mitigation have been published. Departing from the cross-region spread of infection, one study (Kraemer, 2020) investigated the effects of travel restrictions on delaying epidemic outbreaks and reducing the number of cases. This research was followed by a further study on understanding cross-region population flow as a driver of the spatial-temporal distribution of COVID-19 in China (Jia et al., 2020). Through the use of mathematical models on metapopulations, such as the global epidemic and mobility model (GLEAM) (Balcan, 2009), risk source model (Jia et al., 2020), and disease contagion models (Chowell et al., 2004; Newman, 2002), these studies explored the mechanisms affecting the large-scale pandemic and demonstrated the role of human behaviors (especially human mobility) in disease transmission, and also examined the effectiveness of non-pharmacological interventions in virus spread control. With the understanding of the underlying mechanisms influencing epidemic spread and ubiquitous population movement data, data-driven methods and tools, such as deep neural networks (Kapoor, 2020; Ramchandani et al., 2020), have been adopted to predict the spread of infections across states and

* Corresponding authors.
E-mail addresses: chfan@tamu.edu (C. Fan), amostafavi@civil.tamu.edu (A. Mostafavi).

https://doi.org/10.1016/j.cities.2022.103805
Received 9 April 2021; Received in revised form 29 November 2021; Accepted 2 June 2022
Available online 8 June 2022

counties. The well-performed predictions obtained through the use of these methods expand the capability of forecasting future trajectories of increases in the number of cases across different regions (Chang et al., 2020).

Despite enormous progress (Holtz, 2020) in slowing down the cross-regional spread of the pandemic, the COVID-19 pandemic has already entered the phase of community spread. The continuing infection growth accentuates the need for further localized containment measures (Schlosser, 2020). Existing meta-population-based analyses, however, are limited in capturing high-resolution person-to-person transmission to inform the evaluation of localized measures (Balcan, 2009). Recent studies have argued that anonymized mobile phone data, when used properly and carefully, could characterize the epidemic dynamics among human contacts during all stages of the pandemic life cycle (Oliver et al., 2020). Mobility data collected from a large number of devices enables time-resolved characterization of population contact patterns, making it possible to probe the mechanisms by which disease is transmitted among the population with precision unattainable by other data sources, such as surveys. Due to the benefits of mobile phone data (Grantz et al., 2020), studies (Liu, 2018) have attempted to simulate the SARS-CoV-2 transmission on synthetic populations derived from mobile phone data in normal situations and to model the impact of individual-based measures, such as testing, contact tracing and household quarantine (Aleta et al., 2020).

Human behaviors, however, are dynamic, as the pandemic evolves. Limited consideration of the dynamics of human contact patterns would undermine the robustness of localized measures of pandemic mitigation. Models that can explain the virus spread based on dynamic human behaviors are especially needed. In addition, after the execution of stay-at-home orders and business activity restriction policies in the United States during March and April 2020, many states and counties allowed reopening of economic activities to relieve the heavy burden on the economy caused by the shutdown (Bonaccorsi, 2020). Although lifting restrictions could help with economic recovery, the subsequent increase in population contact activities drove the pandemic to a new peak (Kaxiras & Neofotistos, 2020). Clearly, an integrative consideration of infections and economic costs in the decision-making process is critical to inform making tradeoff decisions between the number of infections and the economic costs of containment policies. The existing models do not enable examining this important tradeoff.

In this study, we investigated the dynamic contact patterns from fine-grained, anonymized data for millions of mobile devices along with census and demographic data. Accordingly, we then synthesized weekly contact networks of the populations based on their dynamic contact patterns captured by mobile phone data from the date of first reported cases through the end of June 2020 in ten selected US metropolitan counties. Each agent in the contact networks is associated with a census block group (CBG) within specific residential areas. On top of the dynamic data-driven contact networks, we overlaid a time-varying, degree-based Susceptible-Exposed-Infectious-Recovered (SEIR) contagion model that simulates the case trajectories among the synthetic population. This model allows us to capture the temporal evolution of the pandemic and the spatial distribution of the infections due to the dynamic contact patterns of the population. Through the understanding of the spatial-temporal heterogeneity of disease spread, we could identify the most at-risk populations, quantify the cost and infection trajectories, and provide a quantitative analysis of the effects of combined local containment policies on the costs and infections.

Here, we demonstrate the performance of our model in terms of out-of-sample prediction based on population contact behaviors. The results show multiple waves of virus spread and spatial variation within ten US metropolitan counties. The results also highlight the heterogeneous effects of local policies on contact networks and disease spread across different regions. With these important observations, we propose a method that allows us to quantify the cost of the local policies and recommends optimal combined policies to achieve a better trade-off between economic recovery and virus control.

2. Results

2.1. Dynamic data-driven contact networks

The computational experiments are conducted on large-scale social contact networks. It would be very computationally expensive if we use the actual population to construct the social networks. Hence, we create the agent-based model in which we capture the mobility patterns of the residents from a CBG using mobile phone data and synthesize a smaller group of the population that is proportional to the number of residents in a CBG using census data. To provide a quantitative estimate for the dynamic contact patterns in the population of agents in each county, we first use detailed demographic data from US Census Bureau (Bureau, 2019) to generate a certain number of agents and assign them to specific Census Block Groups (CBGs). During this process, we ensure that the numbers of agents in the CBGs are proportional to the numbers of residents according to the census data. To validate the representativeness of the anonymized mobile phone data, we conducted a correlation analysis to examine the extent to which the number of mobile phone devices is proportional to the number of residents in a CBG from the census data (see in Supplementary Information Section 1). We find that, the correlation coefficient is high enough, indicating that the population of each CBG captured in the mobile phone data is significantly proportional to the population estimated from demographic surveys. In this respect, we consider that the mobile phone data could represent the populations in CBGs in a county, and can capture the mobility of individuals in the population between areas.

Virus transmission occurs through the co-presence, defined as contact through activities, of two or more agents in the same small area. To simulate infection spread among the synthetic population, we build weekly contact networks ($\mathscr{G}$) that encode the dynamic contact patterns of anonymized mobile devices from our empirical data (see *Methods*). The co-presence of two mobile devices in a point of interest (POIs) determines a link between them, and the duration of the co-presence is the weight of the link (ω_{ij}, where i and j are different agents). POIs refer to public common places such as restaurants, museums, nature parks, day care services, gasoline stations. The synthetic contact networks follow the same patterns of degree distribution and duration of contact as what is documented in the empirical data on a weekly basis (Fig. 1).

2.2. Epidemic contagion in contact networks

We implemented a stochastic time-varying and degree-based compartmental model in which the spread of infection relies on the structure of networks and the duration of contacts (Fig. 2a, the state transition diagram). The model is formulated as a system of ordinary differential equations. In this model, we consider that an agent could be in only one of the five states: susceptible ($S(t)$), exposed ($E(t)$), infectious asymptomatic ($I_a(t)$), infectious symptomatic ($I_s(t)$) and removed ($R(t)$) at any time-step t (1d). Ostensibly, the susceptible agents are those who have not been infected at time t. We consider that the susceptible agents have a certain chance of getting infected based on their contact activities with infectious agents (asymptomatic infectious ($I_a(t)$) and symptomatic infectious ($I_s(t)$)) in the contact networks (Centers for Disease Control and Prevention, 2020). Exposed agents ($E(t)$) are those who have been infected but are not yet infectious. That is, the structures of the networks that facilitate the spread of the disease are incorporated, which leads to heterogeneous probabilities of susceptible agents to be infected (Pastor-Satorras & Vespignani, 2001). The agents with higher degrees (more links connecting to them) are more likely to be in contact with infections, and therefore they would be more likely to get infected (Barabási & Pósfai, 2016). Assuming the agents with the same degree behave similarly (Barabási & Pósfai, 2016), we employ degree block approximation to place the agents that have the same degree into the same block

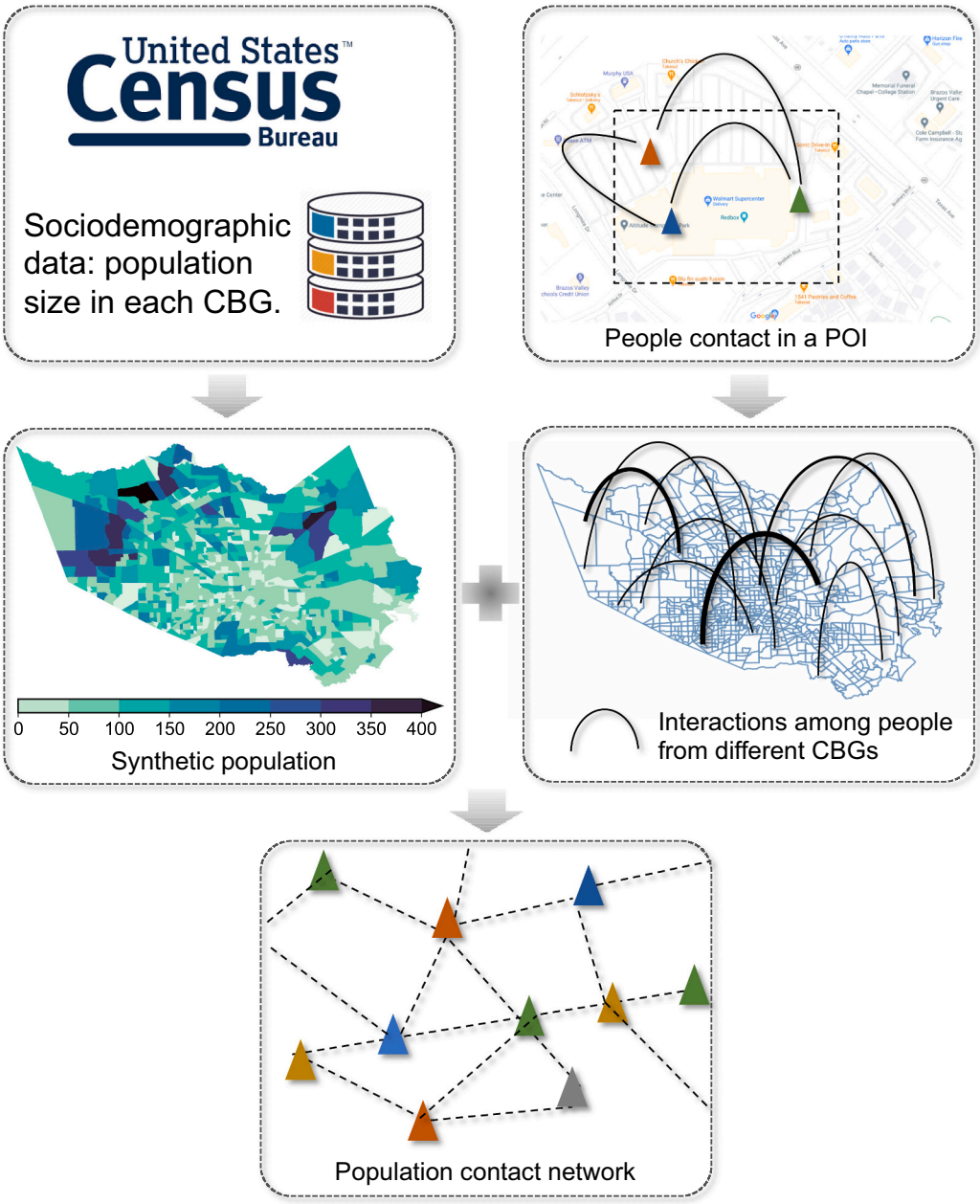


Fig. 1. Schematic illustration of synthesizing the dynamic contact networks from anonymized mobile devices.

(group). This approach allows us to create a separate compartmental model in each group of agents based on their degrees.

Contact networks tend to lack degree correlations, which means the probability that a link points from an agent with degree k to an agent with degree k' is independent of k. Such a heterogenous contact pattern in the group forms the basis to formulate a general differential equation system. The probability that a random chosen link connects an agent in the group of degree k' is the excess degree $k'p_{k'}/\langle k\rangle$, where $p_{k'}$ is the probability of a random chosen agent with degree k' in the contact network. Existing infected agent should have at least one link connecting to another infected agent, the one that transmitted the disease. Therefore, the number of links available for future transmission of the agent is $(k'-1)$. The fraction of infected nodes in the neighborhood of a susceptible agent in a group of degree k is defined as a density function Θ_k. We also consider the agents at the infectious stage can spread the disease and infect other susceptible agents. As such, the density function can be written as:

$$\Theta_k = \frac{\sum_{k'}(k'-1)p_{k'}i_{k'}}{\langle k\rangle} \tag{1}$$

where $i_{k'}$ represents the density of infected agents with degree k' in the network. As observed from Eq. (1), in the absence of degree correlations, the density function Θ_k is independent of k. By differentiating both sides of the equation, we can obtain:

$$\frac{d\Theta_k}{dt} = \sum_k \frac{(k-1)p_k}{\langle k\rangle}\frac{di_k}{dt} \tag{2}$$

Exposed agents are transformed from susceptible agents into infectious at a rate of $1/\alpha$. Hence, the change of the density of exposed agents depends on two transitions: the number of agents transitioned from susceptible; and the number of agents transitioned to the infectious. We formalize the changing rate of the density of exposed agents as:

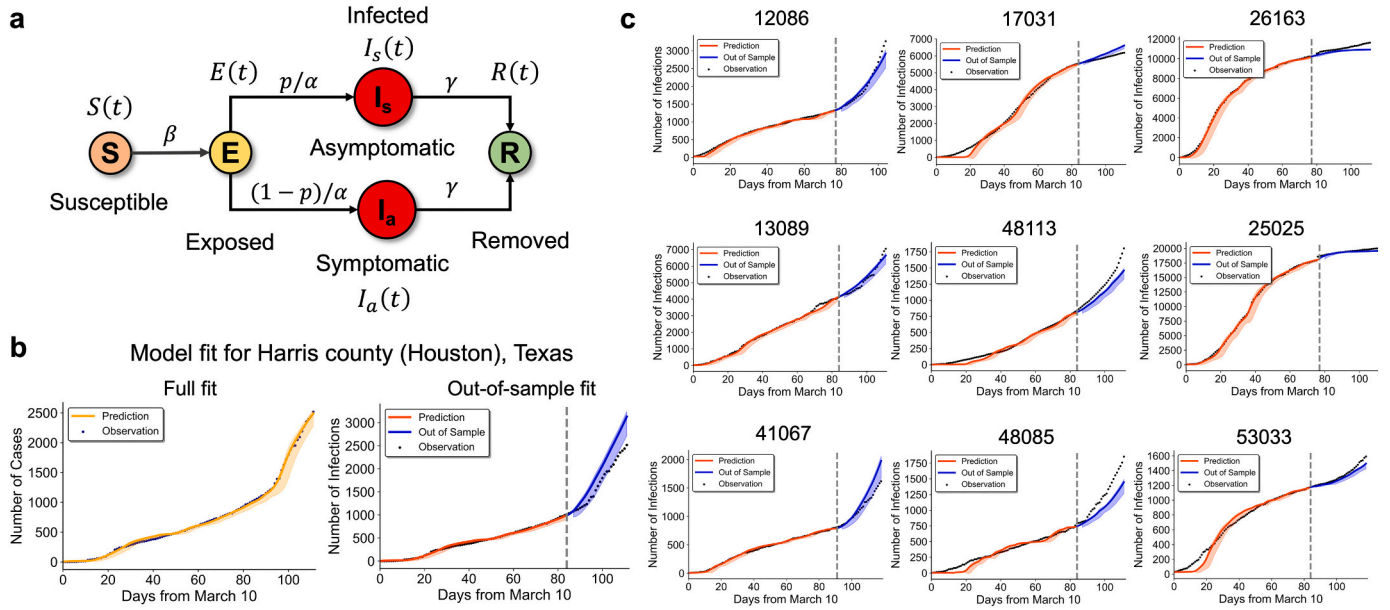


Fig. 2. Model description and fit. a, The state transition diagram illustrating the compartmental model used to capture the transmission process of the SARS-CoV-2 among the population in different states. Specifically, we consider the susceptible (S), exposed (E), infectious asymptomatic (I_a) and infectious symptomatic (I_s). The rates, β, p/α, $(1-p)/\alpha$, and γ indicate the probability of people transformed from the current state to the next state. Here, p is equal to 20 %. More details for the model are provided in the main text and the supplementary information. b, Model fit for Harris County (Houston area), Texas. The plot on the left panel is the full fit on the full range of data, while the plot on the right panel is the out-of-sample prediction. In the out-of-sample prediction, we calibrated the model on the data before 80 days since March 10 and predicted the infections afterwards. c, The model fit and out-of-sample prediction for additional nine US metropolitan counties. The numbers on top of each subplot show the Federal Information Processing Standard (FIPS) codes of the selected counties.

$$\frac{de_k(t)}{dt} = \beta k \cdot \Theta_k(t) \cdot s_k(t) - \frac{e_k(t)}{\alpha} \tag{3}$$

where $s_k(t)$ represents the density of susceptible agents with degree k in the contact network at time t, $e_{k'}$ represents the density of exposed agents with degree k', and α is an inverse of a rate, indicating the mean latent period for the disease. To be consistent with reported results in the literature (Guan et al., 2020), we consider that the SARS-CoV-2 has a latent period of 14 days in this study (α equals 14). The infection rate is proportional to the general infection rate β in the entire contact network and the degree k, which specifies the chance of a susceptible agent with degree k being infected.

Exposed agents will further move to infectious stage, leading to the change of density of infected agents in the network. The change of infected agents also depends on two transitions: one is agents transitioning from the exposed state, and the other is the agents transitioning to recovery status. Therefore, the changing rate of the density of infected agents at time t could be given as:

$$\frac{di_k(t)}{dt} = \frac{e_k(t)}{\alpha} - \gamma \cdot i_k(t) \tag{4}$$

where γ is the recovery rate, meaning that both asymptomatic and symptomatic infections would be removed after a mean infectious period $1/\gamma$. The infections with degree k are composed of asymptomatic infectious and symptomatic infectious agents with degree k. Hence, the density of infections in the contact network $\mathscr{G}$ could be represented as:

$$i_k(t) = i_{ak}(t) + i_{sk}(t) \tag{5}$$

where $i_{ak}(t)$ and $i_{sk}(t)$ represent the density of asymptomatic infectious and symptomatic infectious agents with degree k in the network respectively. Based on the settings in an existing study (Aleta, 2020), we consider that 25 % of infections are asymptomatic infectious and the rest are symptomatic. This study is also aware of the case detection rate, which may vary with prevalence, testing capacity, testing protocols, and reporting fatigue. These factors may in turn all vary spatially and temporally. To take the influences of these factors into account, we calibrate the model in each week so that the model can capture the dynamic infectious situation well.

Then, with a probability of p, exposed agents in the latent state will move to the symptomatic infectious state; otherwise, they will move to asymptomatic infectious. Since the probability p in our model is fixed, we still use $I_k(t)$ in the mathematical formulation of the problem. The probability p, however, influences the trajectory of disease spread in contact networks, which will be discussed later.

Plugging both Eqs. (3) and (4) into Eq. (2), we have:

$$\frac{d\Theta_k(t)}{dt} = \sum_k \frac{(k-1)p_k}{\langle k \rangle} \left(\beta k \cdot \Theta_k(t) \cdot s_k(t) - \gamma \cdot i_k(t)\right) \tag{6}$$

Here, we keep only the first order terms, which means that the $s_k(t)$ could be ignored above, as for small t, $e_k(t)$ and $i_k(t)$ are much smaller than one and $s_k(t)$ is much close to 1. Then, the Eq. (6) could be simplified as:

$$\frac{d\Theta_k(t)}{dt} = \left(\beta \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} - \gamma\right) \cdot \Theta_k(t) \tag{7}$$

To solve this equation, we get:

$$\Theta_k(t) = Ce^{t/\tau} \tag{8}$$

$$\tau = \frac{\langle k \rangle}{\beta \langle k^2 \rangle - \langle k \rangle (\beta + \gamma)} \tag{9}$$

where τ is the characteristic time for the model. Using the initial condition, $\Theta_k(t=0) = C$. Hence,

$$C = \frac{\langle k \rangle - 1}{\langle k \rangle}(e_0 + i_0) \tag{10}$$

where e_0 and i_0 are initial values for the densities of exposed agents and infected agents.

The susceptible agents can only transform to exposed agents. Hence,

we formulate the change of the density of susceptible agents as:

$$\frac{ds_k(t)}{dt} = \beta k \bullet \Theta_k(t) \bullet s_k(t) \tag{11}$$

The removed agents include both agents recovered from the infectious state and agents who died of the disease. Therefore, we determine the changing rate of removed agents using:

$$\frac{dr_k(t)}{dt} = \gamma i_k(t) \tag{12}$$

$$N_k/N = s_k(t) + e_k(t) + i_k(t) + r_k(t) \tag{13}$$

where $r_k(t)$ is the density of removed agents with degree k in the network. The ordinary differential equation is formed in each group of agents with specific degrees. The total number of infected agents is the sum of all infected degree-k agents: $I = \sum_k p_k I_k$, where p_k is the probability density of the agents with a degree of k. Hence, the equations above capture, with a set of equations for all degree-k agents, the time-dependent behavior of the whole system.

With the estimates of populations in each state, the next critical step is to determine which agent is in what state to capture both temporal and geographical patterns of the pandemic spread in the contact network. Here, we adopt the idea of network percolation in three transitions to simulate the state transition of specific agents. First, susceptible agents are more likely to get infected if they are exposed to (contact) infections for a long time. Hence, to select the agents who will make transition from susceptible state to exposed state in the next time step, we compute the time that susceptible agents spent in contact with infectious agents. The probability of a susceptible agent being infected ($p_i^{(S\to E)}$) is proportional to contact time with infectious agents.

$$p_i^{(S\to E)} \propto \left(\sum_{j\in I} \omega_{ij} \right) \tag{14}$$

where i is a susceptible agent and j is one of the infectious agents.

Second, exposed agents in latent state have two directions for transitioning to the next state: symptomatic and asymptomatic infections. We assume that symptomatic infections tend to have fewer movements and contact activities, while asymptomatic infections may maintain their behaviors as normal. Based on these assumptions, we compute the differences in contact activities and how the exposed agents behave during the following week and the current week. The probability of an exposed agent having symptomatic infection is proportional to the differences in the contact activities. The more reduced the contact of an exposed agent in the following week, the higher the probability of transitioning to a symptomatic infectious state (rather than an asymptomatic infectious state). We denote the probability of transitioning to symptomatic infection ($p_i^{(E\to I)}$) as follows:

$$p_i^{(E\to I_s)} \propto \left(\sum_j \omega_{ij}^{(T+1)} - \sum_j \omega_{ij}^{(T)} \right) \tag{15}$$

where, $\omega_{ij}^{(T+1)}$ is the contact time between i and j on week $T+1$.

Third, both asymptomatic and symptomatic infectious agents would transition to removed agents. They may recover from the disease or die of it. The mean duration of infection is about 14 days, based on the information from the Centers for Disease Control and Prevention (Healthcare workers, n.d.). Hence, we consider the probability of an infectious agent being removed in the next time step follows a Gaussian distribution with a mean of 14 and a standard deviation of 3.

$$p_i^{(I\to R)} \sim \mathcal{N}(14, 3) \tag{16}$$

Each agent in the state of infection would be assigned a probability at each time step, and the agents with a higher probability are more likely to be selected for state transition.

Through the above-described contagion and percolation process, this model allows us to capture both temporal evolution and geographical distribution of the infections and how the epidemic spreads in human networks due to contact activities. To obtain model parameters over time, we calibrated the model for each week and in every county in the United States using cumulative confirmed cases obtained from the *The New York Times* (2020).

2.3. Model fitting and validation

To validate the model, we showed the performance of the model in fitting real case data and predicting out-of-sample cases using the model calibrated on the period before. Specifically, we split the study period into two parts: the first 12 weeks as the learning set for fitting and calibrating the model; the balance of the time (4 weeks) as the testing set for predicting the confirmed cases. We first fit the model taking the weekly contact networks in the first 12 weeks as the input and obtained the values of model parameters. Due to the delay in case testing and reporting, our model implemented on the contact network in any week was fitted on the cases in the following week by minimizing the root mean square error (RMSE). Hence, we define the loss function as:

$$\min_{\beta,\gamma} \mathscr{L}_{\mathscr{E}} = \sqrt{\sum_{d=1}^{D} \frac{(\hat{y}_d - y_d)^2}{D}} \quad \text{s.t.} (\hat{y}_d - y_d)^2 \le 0.01 y_d \tag{17}$$

where, d is the day in the fitting process, D is the number of days in total, $\hat{y}_d$ is the predicted number of cases, and y_d is the actual reported number of cases. It should be noted that the constraints we added here were to ensure that the number of cases on the last day of the prediction was close to the actual reported cases. That is, the deviation of the predicted results would not influence the prediction of the following weeks.

It is evident that using mathematical derivation to solve the problem of Eq. (17) and get analytical solutions is challenging. That is because the functions are usually not continuous and differentiable. A more efficient way is to calculate the numerical solutions through a heuristic algorithm. In this study, we employed the global pattern search algorithm as a derivative-free numerical optimization method to identify the optimal point which can minimize the loss and satisfy the constraint. The estimated values for model parameters and the optimized RMSE are shown in the supplementary information.

Through the training process, we obtained the model parameters including β and γ. The stochasticity is introduced in the model through the initialization of the exposed and infected agents as well as the specific infected agents in the groups of degrees. (See supplementary information for more details.) Running the resulting models on contact networks in the testing set, we predicted the confirmed cases in the testing weeks. Specifically, the parameters obtained from the most recent week in the training set for each county are used to make the predictions of out of sample data (testing set). We evaluated the predictive performance of the model by comparing the differences between the predicted cases and actual reported cases. Sub-figures in Fig. 2b and c show that the model fits the out-of-sample case data very well, demonstrating that this model is effective in extrapolating beyond the training set to future periods.

2.4. Spatial-temporal heterogeneity of infections

Since the synthetic population of the contact networks is generated based on US Census data together with the results of the model, we can characterize the differential spread of SARS-CoV-2 across the CBGs in a county. Through examining the spatial and temporal mechanisms of the disease spread, we are able to quantify different levels of infection in different places and provide a quantitative approach for modeling the execution of local containment and recovery policies.

We explain an example of the COVID-19 pandemic in Harris County

(Houston area), Texas. Fig. 3a and d show the predicted numbers of agents in exposed, removed, and symptomatic infectious states by fitting the model with reported cases. We observed that the pandemic situation evolved over time. For example, there are multiple peaks of infections in the study period. This result indicates that our model is distinct from existing simulation models and could capture realistic patterns of virus spread over time. Due to the lack of dynamic fined-grained contact networks, existing simulation models tend to simulate the spread of a disease on a static contact network in which the connections among the agents are not time-varying. Such simulations may not be able to account for the evolving human behaviors and explain the temporal variations and spatial heterogeneity of infections.

Fig. 3b shows an example of the geographical distribution of infections across Harris County. It is evident that the infections are geographically heterogeneous. This finding explains the contribution of contact patterns of people to the infection rates of their CBGs. People who engage in intensive contact activities cause their CBG to be intensively infected. The geographical heterogeneity in number of cases is further demonstrated in other counties (Fig. 3e). The mean entropy of the infection distribution grows rapidly from 0.5 in the first week of March 2020 to around 1.5 in late March 2020, and it remains at a high level until the end of the study period. This result demonstrates that the infection patterns are extremely heterogeneous across different areas of the county and implementing uniform containment or reopening policies may lead to severe burdens on the economy or infection reduction efforts. Our model shows an important capability for uncovering the mechanism of disease spread, which could be used in formulating local policies.

The results above raise an important need for localized policies, such as geo-fencing in high-infection CBGs and reopening in low-infection CBGs. Since policy development is based on the pandemic situation at the moment, the evaluation of policy effectiveness requires projection to the future. That is, the containment policies should apply to areas that show high infection at the current time step and will maintain a high level of infection in the future in the absence of containment policies. By doing so, the localized containment policies would have an optimal outcome in containing the spread of the disease. In response to this question, we split the range of infection percentages of the CBGs in a county into five levels with equal ranges. The example for Harris County is shown in in Fig. 3c. We find that the numbers of cases across the CBGs are highly unequally distributed in these five levels. Most of the CBGs are at the lowest infection level, while a few CBGs are greatly infected. Then, we examine the level of infections of these CBGs in different months (Fig. 3f). We observe little variances regarding the infection levels of the March CBGs in each group. This result demonstrates that the CBGs maintain their levels of infection during the pandemic, implying that adopting targeted policies for the CBGs with high infection levels at this time step is essential to preventing further deterioration of these CBGs. This finding provides important evidence for developing localized policies, which will be discussed in the following section.

2.5. *Effects of local policies on network structure and daily infections*

With the unveiled spatial-temporal patterns of the pandemic, local policies should account for such heterogeneous impacts. Since the outbreak began in early March 2020, governments and policymakers have introduced large-scale restriction measures, such as mobility control and mask use. Typical measures in mobility control include area lockdown and stay-at-home orders. The restriction of human contact activities, however, induced a heavy burden on the local economy, especially on labor-intensive businesses. To mitigate the economic pressures, many counties and states lifted their restrictions and allowed local businesses to reopen. These above-mentioned local policies were commonly adopted in the majority of counties in the United States. The local policies and their effects on contact networks lead to multiple peaks in the pandemic. This situation prompts a question regarding the extent to which these policies would reduce the pandemic spread.

For these reasons, we project the direct impacts of typical local policies on human contact networks and estimate the consequences of contact network changes on infections. Three different scenarios through the implementations of three local policies are summarized here:

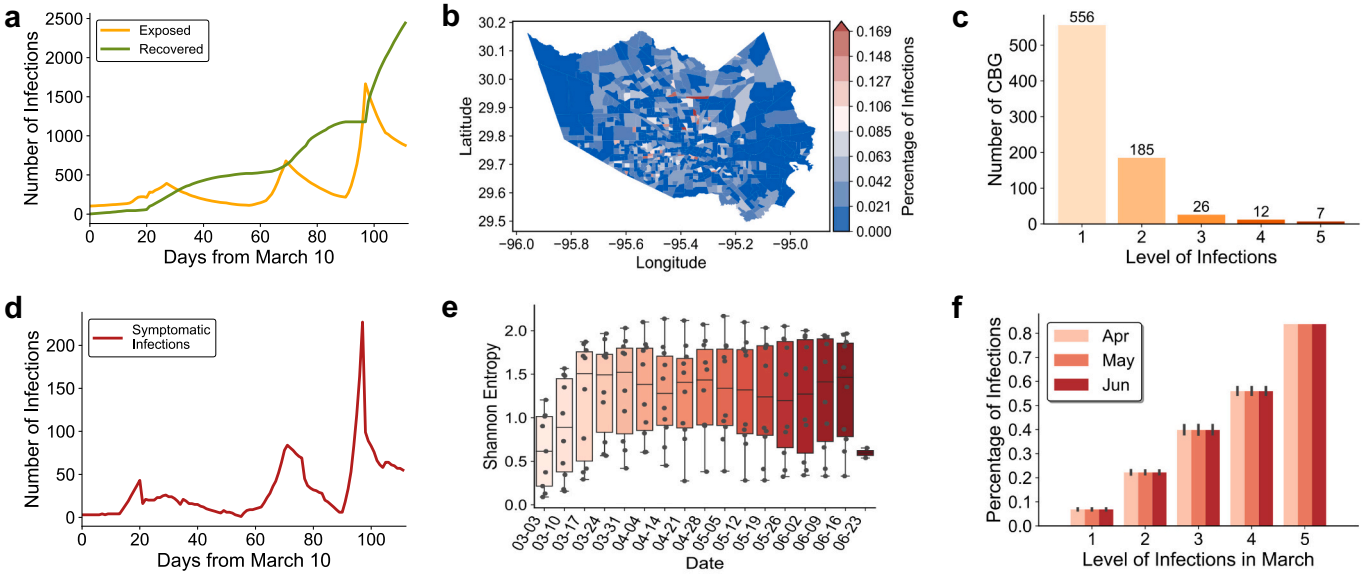


Fig. 3. The outcomes of disease spread in the data-driven contact networks. a, Daily number of exposed infections and cumulative number of recovered infections in Harris County, Texas, simulated by our model. b, Spatial distribution of infections characterized by the percentage of people who were infected on the 80th day since March 10 across CBGs in Harris County, Texas, simulated by our model. Spatial distributions of infections in other counties are provided in supplementary information. c, We categorize the CBGs in Harris County into five levels of infections. The range of percentages of infections at each level is the same. d, Daily new symptomatic infections in Harris County, Texas, simulated by our model. e, The results of Shannon Entropy quantifying the spatial heterogeneity of infections across CBGs in each county for the ten studied counties. f, The percentage of infections in April, May, and June in CBGs at levels of infections determined in March.

- *Mobility control* (*CBG lockdown*) *scenario*: Mobility control measures including business lockdown and stay-at-home orders with the intent of limiting the number of contacts a person would have. In particular, local containment policies are developed to contain the activities of people from specific areas. Together with the findings from the last section, we rank the CBGs based on their percentage of infections from high to low. Then the mobility control measures are simulated by removing the connections of agents belonging to specific CBGs. We started containing the activities of agents from CBGs with a high level of infections, and selected 0 %, 25 %, 50 %, 75 % and 100 % of CBGs to be contained.
- *Mask use scenario*: The SARS-CoV-2 infection is transmitted predominately by respiratory droplets generated when people talk to each other (Moghadas et al., 2020; Worby & Chang, 2020). The longer the face-to-face communication, the higher the chance of infection. Masks are primarily intended to reduce the emission of virus-laden droplets and reduce inhalation of these droplets by the wearer (Fischer et al., 2020). Hence, we simulate the mask use scenario by reducing the weights of the links between the agents with masks. If the weights after the reduction of a specific value become negative, we consider there is no link that would allow a disease to be transmitted. As such, with a certain percentage of agents wearing masks, the contact network would be sparser, implying a lower probability that a disease can transmit across the networks. It should be noted that the selection of agents wearing masks is random, without consideration of their CBGs and infection levels.
- *Reopening scenario*: When reopening policies are enacted, businesses reopen and population activity increases. Reopening policies allow people to re-engage in contact activities to the level of normal conditions (Vermund & Pitzer, 2020). Hence, to simulate the scenario of reopening, we extracted the contact networks in February before the outbreak of the pandemic and synthesized connections similar to that of normal conditions for selected agents. The localized reopening policy is also executed in the CBGs in terms of their levels of infections. That is, we start reopening CBGs with the lowest level of infections then selecting CBGs with increasingly higher levels of infection based on specified reopening rate (percentage of reopened CBGs among all CBGs). In this study, the reopening rates are 0 %, 25 %, 50 %, 75 % and 100 %.

Fig. 4 shows the effects of local policies on the structure of contact networks and the consequences of network changes on disease spread in Harris County. (See results for other counties in the supplementary information.) The first panel of the figure shows the effects of mobility control measures. As shown in Fig. 4d, the total weights of the contact networks in Harris County decreased 50 % in about one month since March 10, 2020 and remained stable until the end of the study period. Mobility control measures implemented on the 13th week since March 10 significantly reduced the total weighting in the network. The reduction of the weighting is almost proportional to the percentage of the controlled CBGs in a county. When all CBGs are contained, the total weights of the network are 0, meaning that no links exist present in the contact network. This would be an ideal outcome of local containment, which cuts all possible transmission trajectories of the disease. Hence, mobility control policies could lead to promising results in terms of pandemic containment (Fig. 4a). All curves with mobility control policies indicate a peak of infection one or two weeks after the policy implementation, as is shown in historical data. That is because the frequent contact activities enable a denser contact network, while the total weight of the network does not change dramatically. This conclusion is realistic as it indicates the effect of contact activities in disease spread. With the increase of controlled mobility in CBGs, the magnitude of the peak infections decreases. When 75 % or more CBGs are controlled for mobility, the peak number of infections is at a similar level as or even a lower level than the previous peak. This result implies that an aggressive mobility control policy is effective and necessary to change the trajectory of the pandemic, or at least to prevent the pandemic from worsening.

The middle panel of Fig. 4 shows the effects and consequences of mask use on the pandemic spread. When agents are in contact with others, masks reduce infection to a certain degree. Hence, the use of masks does not have evident effects (Fig. 4d), as mobility control

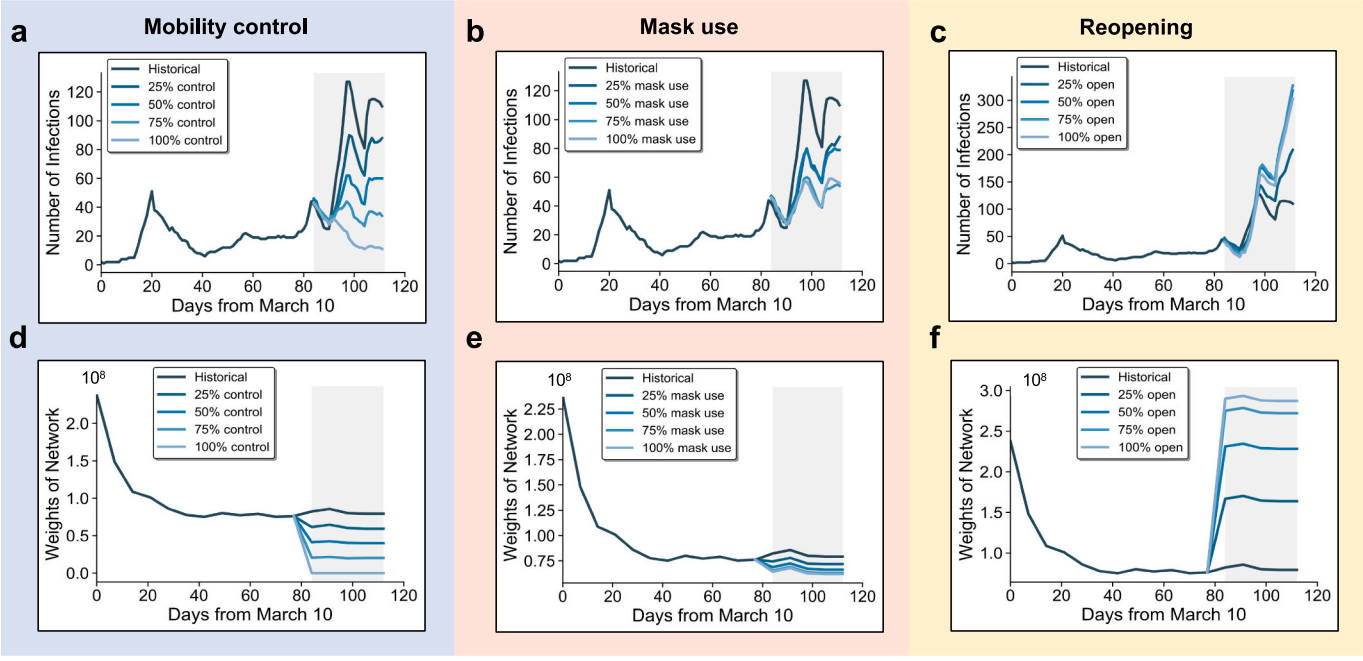


Fig. 4. Effects of localized policies on daily infections and network structure for a variety of policy implementation levels. Each panel represents effectiveness in terms of relative levels of policy implementation on number of daily infections and the contact network. The impacts of varying policy levels on daily new infections a–c are shown for different policies: mobility control, mask use, and reopening. The outcomes of varying policy levels on the weighted contact networks d–f are shown with different policies. This study selects different levels of 0 %, 25 %, 50 %, 75 % and 100 % for policy implementation on the synthetic population.

policies explicitly cut the connections among the agents and break down the contact network. Mask use is still essential since it can to some degree moderate the pandemic (Fig. 4b). An increase in the number of people wearing masks leads to a significant decrease in the daily new infections and the magnitude of the peak number of infections. In an extreme case in which 90 % agents wear masks, the pandemic would be still out-of-control. That is because, there are some agents who contact others frequently and for a long time and some of them are infectious. The role of wearing masks is weakened in such cases. Hence, simply requiring all people to wear masks is insufficient to contain the epidemic.

The last panel of the figure indicates the outcomes of reopening policies. Although the stay-at-home order has been lifted since late May and June 2020, population mobility and contact activities did not revert to their pre-pandemic patterns. As we simulate the contact activities of some agents based on their regular activities in normal conditions, the contact networks get denser and the weights of the network increase drastically (Fig. 4f). Such change in the network causes a sharp increase of the daily infections, two or three times the historical reported cases. In addition, as the hardest-hit CBGs reopen, the growth of daily infections rises even faster. This result shows a warning that untimely reopening policies during the pandemic could neutralize all previous control efforts.

In summary, we have observed diverse effects of different policies and various levels of implementations of these policies on the structure of contact networks and the pandemic spread. Purely from a perspective of pandemic containment, both mobility control and mask use are necessary and effective in containing the transmission of the disease through population contact behaviors. During pandemic peak and economic recession, however, recovery of economic activities is also of great concern. Despite these exigencies, we must still take into account the larger economic loss which may be caused by a possible worsening pandemic in the future if we simply reopen regions without adequate containment measures (Fan, 2020).

2.6. Impacts of combined policies on costs and infections

The above analyses and results raise two important questions: what combination of policies would be more effective in containing the pandemic; and how we can make a trade-off between costs and infections for policy development. In this section, we present a method to quantify the costs of combined policies and also project the pandemic situation after implementing these combined policies.

In our simulation model, we assume the system is consistent, meaning that no inflow nor outflow of agents occurs during the study period. The total number of agents remains the same throughout the study period. In addition, combined policies are consistently implemented during the test period. Thus, the timing of the policy implementation is not a factor influencing the calculation of the costs and the projection of infections. Based on these assumptions, we consider the costs of the policies to be proportional to the size of the population affected by the policies. Accordingly, we can assign a cost factor to each group of the population under specific policies to roughly quantify the costs of the policies. The cost function is formulated as follows:

$$\mathscr{C} = \mu_1 Pop_m + \mu_2 Pop_c - \mu_3 Pop_r + \mu_4 Pop_i \tag{18}$$

where μ_1, μ_2 and μ_3 are the cost factors for three types of populations under the policies. The values of these cost factors can be selected based on an estimation of the costs in specific local regions. Hence, the cost factors may be variant from region to region. The factors such as hospitalization, hospital capacity, age-dependent severity of illness and adherence with public health measures could also be incorporated in the cost factors to enable an accurate estimation. See more details in the supplementary information. In addition, Pop_m is the percentage of the population wearing masks, Pop_c is the percentage of the population in the CBGs with mobility control, Pop_r is the percentage of the population in reopened CBGs, and Pop_i is the number of new infections during the implementation of the policies. To compute the costs and project the infections for all combinations of the policies, we conducted a test for several levels of policy implementation. Specifically, we selected the values for three policy-related variables (i.e., Pop_m, Pop_c, and Pop_r) from 0.1 to 0.9 with increments of 0.1, estimated the infections under the specified policies, and computed the costs for all combinations of these variables.

To compare costs of different local policies, we focused mainly on relative values of costs, which indicate extra savings or expenses of one policy compared to another. In other words, the cost values are not required to be precise in this analysis since relative costs are sufficient for us to identify cost-effective policies. As long as relative costs are reasonable, the assumption is defensible. Hence, we define μ_1 to be 5, μ_2 to be 30, μ_3 to be 5, and μ_4 to be 1. Taking the example of Harris County, Texas, it is clear that different combinations of local policies result in varying levels of infections in the test period (Fig. 5a). Furthermore, by maintaining the stability of one policy, we can observe the cost and infection changes brought by the changes in the other two policies. For example, as we keep 10 % of reopening in CBGs at the lowest level of infection, a low rate of mobility control and mask use leads to a great number of infections (Fig. 5f). The number of infections decreases as we increase mobility control and mask use. Meanwhile, the containment effect of mobility control is more evident than mask use. Such a pattern of effects by local policies influences the costs. As shown in Fig. 5b, the most cost-effective strategy in this case is to implement mobility control (i.e., geo-fencing) in 10 % of CBGs which are highly infected with an 80 % mask-wearing rate. This result highlights the importance of mask use to achieve pandemic containment. In another case, we consider the rate of mobility control to be stable (10 %) and change the rates of CBGs with reopening and the percentage of people wearing masks. We find that the reopening of CBGs significantly worsens the epidemic situation, leading to an explosion of infections (Fig. 5g). Due to the reopening and associated contact activities, the actual cost of a high rate of reopening is quite low, although a large number of people get infected (Fig. 5c). An optimal strategy would be increasing the percentage of the population wearing masks and reopen the majority of the CBGs that are not highly infected. Finally, we examine the impact of changing the rates of mobility control and reopening and maintaining 10 % of the population wearing masks. Fig. 5h shows a clear pattern that the upper left corner is dark, while the lower right corner is bright, indicating that reopening establishes a number of connections between people and subsequently provides a path for disease spread. Due to the effect of reopening on economic recovery, the cost could be relatively low when 10 % of the hardest-hit CBGs are under mobility control (Fig. 5d).

Counties have different pandemic situations, which require different compound policies for containment. In this step, we apply our method to ten counties to identify their cost-effective combined policies. The optimal combined policy strategies are plotted in Fig. 5e. The cost-effective combined policies for the majority of the counties, such as Dallas County, Texas (FIPS county code 48113, Dallas) and Wayne County, Michigan (FIPS county code 26163, Detroit) need a low percentage of CBG reopening, a high proportion of population wearing masks, and a large number of CBGs under mobility control. These counties tend to be among the hardest hit during the test period. Other counties, such as King County, Washington (53033, Seattle), and Suffolk County, Massachusetts (25025, Boston), require reopening for fewer infected CBGs to recover their economy, but severely infected CBGs should remain under mobility control. These findings indicate that mobility control is not always a fit-to-all option and not always cost-effective. Different counties require specific combined policies for containing the pandemic and to efficiently recover economic activities.

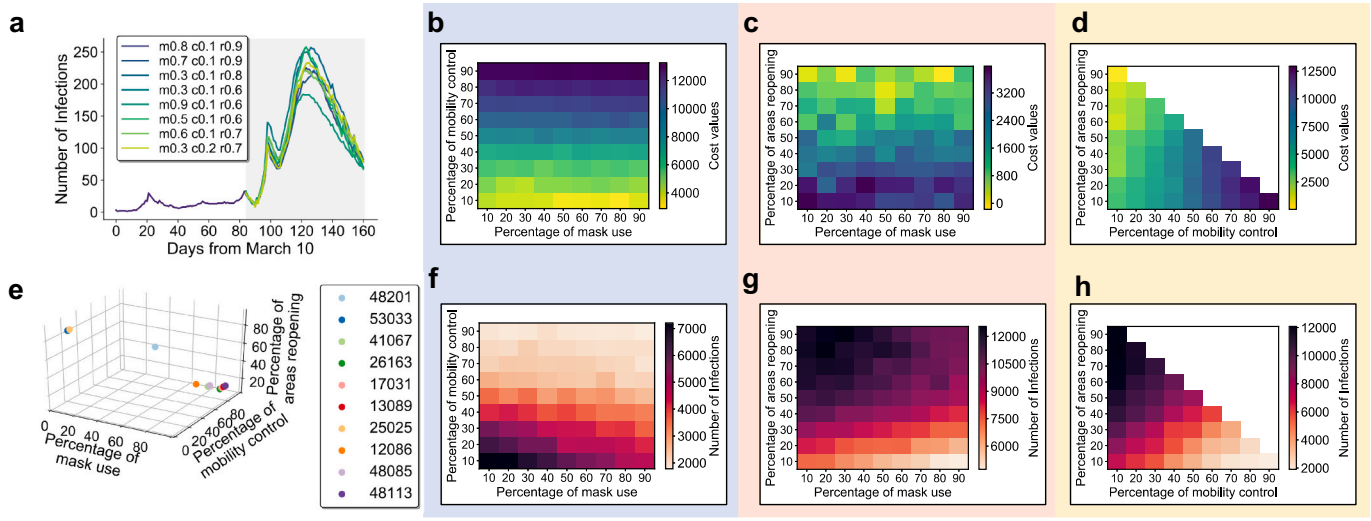


Fig. 5. Outcomes of combined local containment policies and searching for optimal combinations. a, Daily new infections under varying levels of compound policies (selected examples shown) in Harris County (FIPS county code 48201), Texas. To be visually clear, the plot shows only the outcomes for eight combinations of local policies. The estimated cost b-d and the total number of infections f-h with varying levels of compound policies. By keeping the reopen rate at 10 %, the cost and rate infections for varying levels of mobility control and mask use are shown in b and f, respectively. By keeping the mobility control rate to be 10 %, the cost and infection rate for varying levels of reopening and mask use are shown in c and g, respectively. By keeping the rate of mask use to be 10 %, the cost and infection rate for varying levels of reopening and mobility control are shown in d and h. e, The optimal combination of the three local policies for the selected ten counties.

3. Discussion and concluding remarks

We present a data-driven contact network model that incorporates an epidemic contagion process. Using the synthetic contact network built based on anonymized data related to the contact activities from millions of people in ten major infected counties during the COVID-19 pandemic, our model performs well in predicting the spread of the disease in contact networks. Because our simulation is agent-based, we can identify the CBGs of the infected agents and assess the spatial-temporal heterogeneity of the pandemic in a county. This outcome is of particular importance since it addresses the challenge of geographical spread of the disease. As the model suggests, cross-CBG contact activities intensify disease transmission and enable a prediction of the infection distribution, which is consistent with the results of existing studies at national (Jia et al., 2020) and international (Wells, 2020) scales. The findings regarding the spatial-temporal patterns of the pandemic suggest localized containment strategies and inform us in which fraction of the population of urban areas the policies should be implemented. Our study can inform policymakers seeking to adopt area-specific policies instead of county-wide uniform measures (Chen et al., 2020).

With the prevalence of mobility control, mask use and reopening policies (Schünemann et al., 2020), we examined the impacts of these policies on the structure of contact networks and the consequences of structural changes on epidemic spread. We find that mobility control is one of the most effective measures in containing the pandemic, yet imposes a heavy burden on the economy and causes onerous societal disruptions. This conundrum raises an urgent need for combined policies to balance the trade-off between economic aspects and infection cases. Our results show that the cost-effective strategy varies across different counties based on the local epidemic situation. Large-scale mobility control is not fit-to-all. An appropriate composition of the policies may achieve a relatively low infection level and also a low cost to society in the case of partial reopening and economic recovery. Our results and findings could be informative for government agencies for local containment strategies and for public response planning. For example, policymakers may increase testing and quarantine in highly infected CBGs, while lifting control for less infected CBGs.

Our study also confronts some limitations, specifically in the data set we use. It should be noted that the data does not cover the contact activities of all populations. For example, the activities from people who do not opt-in to the data-sharing contract of the data provider cannot be captured. Contact activities that occurred in non-points of interest may not be included in building the contact links in the network. These limitations in the dataset notwithstanding, fine-grained mobile phone data is widely adopted in modeling epidemics and informing public health policies (Oliver et al., 2020). In addition, although our model can accurately predict the pandemic using contact networks, the results do not imply the exact causal relation between contact activities and epidemic spread. Further studies are needed in improving the model to infer the in-depth mechanisms of epidemic spread and to substantiate effective policies in pandemic containment.

4. Materials and methods

4.1. Data sets

We use geolocation data provided by Veraset, Inc. (Veraset, n.d.), a data company that collects anonymized location data from numerous applications and from a large number of devices. Veraset, Inc. collects 2 to 3 billion location event data points every day from >30 M unique devices across the United States. The anonymized devices opted-in to provide access to their location data through thousands of apps and software development kits (SDKs) with Veraset. The data was shared under a strict contract with Veraset through their collaborative program in which they provide access to de-identified and privacy-enhanced mobility data for academic research only. In this study, we filter the data for ten of the most infected counties from January 1 through June 27, 2020. The data contains anonymized device IDs, timestamps, and precise geographical coordinates of dwelling points. We legally performed the analysis and were required never to attempt to identify any individuals.

To identify the contact activities at points of interest, we employed the geolocations of the POIs from SafeGraph (n.d.), a data company that has documented the geographical information and business information about physical places in the United States.

Demographic data at the census block group level are adopted from the US Census Bureau's American Community Survey (ACS) (United States Census Bureau, 2019). The population size of each CBG which we

use to synthesize agents in our model is the most recent one-year estimate (2018).

We calibrated our model using cumulative reported cases from the data dataset published by the *The New York Times* (2020). The dataset documents cumulative COVID-19 infections every day for all the counties in the United States. We fit out the model on these cumulative daily data to acquire the values of model parameters.

4.2. Data-driven contact network

The model considers a weighted network $\mathscr{G} = (\mathscr{V}, \mathscr{E}_t, \mathscr{W}_t)$ with stable agents $\mathscr{V}$, time-varying links $\mathscr{E}_t$, and link weights $\mathscr{W}_t$. We define the agents $\mathscr{V}$ and assigned them to specific CBGs based on the population size of each CBG documented in the ACS demographic data. Then using the dwelling locations of the devices provided in the Veraset data, we considered two devices to be in contact if they presented in the same POI at the same time. Here, the dwelling locations of the devices are defined according to the time the devices spent in a location. We use 5 min as the threshold amount of time to filter out the locations where the devices might be waiting for traffic lights. The length of the overlap of their time spent in the POI is denoted as the duration of their contact. As such, we create an empirical weighted contact network among the devices in the Veraset data. In addition, we estimated the home CBGs of the devices. The links between any pair of CBGs can be extracted. Based on the proportion between the population size of empirical data and the synthetic population, we have a scalar to quantify the difference between the number of devices and synthetic agents. The scalar is then applied to determine the number of links that should be synthesized for each pair of CBGs. The weights of these links follow the distribution of the weights in the empirical contact network. Next, we assign the links to the synthetic agents in these two CBGs, where the degree distribution for the agents is consistent with the degree distribution of the devices in the same CBGs in the empirical data. We generate and assign links by repeating the above steps for all pairs of CBGs. This approach allows for creating synthetic contact networks based on actual mobility data and to maintain key structural and attribute information of the agents and the networks.

We further conducted a test using Kullback-Leibler Divergence to ensure the generated contact networks remained the structural properties of the corresponding empirical network. The detailed results suggest that the degree distributions of the generated contact networks are similar to the degree distributions of the empirical networks (supplementary information). That is, our data-driven contact network is reliable for modeling realistic human contact patterns and pandemic spread.

Data availability

The data that support the findings of this study are available from Veraset Inc., but restrictions apply to the availability of these data, which were used under license for the current study. The data can be accessed upon request submitted on veraset.com. Other data we use in this study are all publicly available.

Code availability

The code that supports the findings of this study is available from the corresponding author upon request.

Computing interests

The authors declare that they have no computing interests.

CRediT authorship contribution statement

C.F., X.J., and A.M. designed the study. X.J., C.F., and R.L. implemented the method and empirical case study. A.M. support with data acquisition. C.F. and A.M. wrote the main manuscript. All authors reviewed the manuscript.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgement

This material is based in part upon work supported by the National Science Foundation under Grant SES-2026814 (RAPID), the National Academies' Gulf Research Program Early-Career Research Fellowship, the Amazon Web Services (AWS) Machine Learning Award, and the Microsoft AI for Public Health Grant. The authors also would like to acknowledge the data support from Veraset Inc. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, Amazon Web Services, Microsoft, SafeGraph or Veraset, Inc.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cities.2022.103805.

References

Aleta, A. (2020). *Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic*. medRxiv. https://doi.org/10.1101/2020.05.06.20092841

Aleta, A., et al. (2020). Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nature Human Behaviour, 4*, 964–971.

Auray, S., & Eyquem, A. (2020). The macroeconomic effects of lockdown policies. *Journal of Public Economics, 190*, Article 104260.

Balcan, D. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. Available at: *Proceedings of the National Academy of Sciences, 106*.

Barabási, A.-L., & Pósfai, M. (2016). *Network science*. Cambridge University Press.

Bonaccorsi, G. (2020). Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences, 117*, 15530 LP – 15535.

Bureau, U. C. (2019). National and state population estimates. Available at: https://www.census.gov/newsroom/press-kits/2019/national-state-estimates.html. (Accessed 6 October 2020).

Centers for Disease Control and Prevention. (2020). Scientific brief: SARS-CoV-2 and potential airborne transmission. Available at: https://www.cdc.gov/coronavirus/2019-ncov/more/scientific-brief-sars-cov-2.html. (Accessed 11 July 2020).

Chang, S., et al. (2020). Mobility network models of COVID-19 explain inequities and inform reopening. *Nature*. https://doi.org/10.1038/s41586-020-2923-3

Chen, P., Wu, K., & Ghattas, O. (2020). *Bayesian inference of heterogeneous epidemic models: Application to COVID-19 spread accounting for long-term care facilities*.

Chinazzi, M. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science (80-.), 368*, 395 LP – 400.

Chowell, G., Hengartner, N. W., Castillo-Chavez, C., Fenimore, P. W., & Hyman, J. M. (2004). The basic reproductive number of ebola and the effects of public health measures: The cases of Congo and Uganda. *Journal of Theoretical Biology, 229*, 119–126.

Fan, C. (2020). *Effects of population co-location reduction on cross-county transmission risk of COVID-19 in the United States*.

Fischer, E. P., et al. (2020). Low-cost measurement of face mask efficacy for filtering expelled droplets during speech. *Science Advances, 6*.

Grantz, K. H., et al. (2020). The use of mobile phone data to inform analysis of COVID-19 pandemic epidemiology. *Nature Communications, 11*, 4961.

Guan, W., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *The New England Journal of Medicine, 382*, 1708–1720.

Headey, D., et al. (2020). Impacts of COVID-19 on childhood malnutrition and nutrition-related mortality. *Lancet, 396*, 519–521.

Healthcare workers . Duration of isolation and precautions for adults with COVID-19. The Centers for Disease Control and Prevention Available at: https://www.cdc.gov/coronavirus/2019-ncov/hcp/duration-isolation.html. (Accessed: 14th November 2020).

Holtz, D. (2020). Interdependence and the cost of uncoordinated responses to COVID-19. *Proceedings of the National Academy of Sciences, 117*, 19837 LP – 19843.

Jia, J. S., et al. (2020). Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature*. https://doi.org/10.1038/s41586-020-2284-y

John Hopkins University. (2020). CSSEGISandData/COVID-19: Novel Coronavirus (COVID-19) cases. Github. Available at: https://github.com/CSSEGISandData/COVID-19. (Accessed 18 May 2020).

Kapoor, A. (2020). *Examining COVID-19 forecasting using spatio-temporal graph neural networks*.

Kaxiras, E., & Neofotistos, G. (2020). Multiple epidemic wave model of the COVID-19 pandemic: Modeling study. *Journal of Medical Internet Research, 22*. e20912-e20912.

Kraemer, M. U. G. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science (80-.), 368*, 493 LP – 497.

Liu, Q.-H. (2018). Measurability of the epidemic reproduction number in data-driven contact networks. *Proceedings of the National Academy of Sciences, 115*, 12680 LP – 12685.

Moghadas, S. M., et al. (2020). The implications of silent transmission for the control of COVID-19 outbreaks. *Proceedings of the National Academy of Sciences of the United States of America, 117*, 17513–17515.

Newman, M. E. J. (2002). Spread of epidemic disease on networks. *Physical Review E*. https://doi.org/10.1103/PhysRevE.66.016128

Oliver, N., et al. (2020). Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Science Advances.* , Article eabc0764. https://doi.org/10.1126/sciadv.abc0764

Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic dynamics and endemic states in complex networks. *Physical Review E, 63*, 66117.

Polyakova, M., Kocks, G., Udalova, V., & Finkelstein, A. (2020). Initial economic damage from the COVID-19 pandemic in the United States is more widespread across ages and geographies than initial mortality impacts. *Proceedings of the National Academy of Sciences.* , Article 202014279. https://doi.org/10.1073/pnas.2014279117

Rader, B., et al. (2020). Crowding and the shape of COVID-19 epidemics. *Nature Medicine*. https://doi.org/10.1038/s41591-020-1104-0

Ramchandani, A., Fan, C., & Mostafavi, A. (2020). DeepCOVIDNet: An interpretable deep learning model for predictive surveillance of COVID-19 using heterogeneous features and their interactions. *IEEE Access, 8*, 159915–159930.

SafeGraph. Places data & foot-traffic insights. Available at: https://www.safegraph.com/. (Accessed 18 November 2020).

Schlosser, F. (2020). *COVID-19 lockdown induces disease-mitigating structural changes in mobility networks*.

Schünemann, H. J., et al. (2020). Use of facemasks during the COVID-19 pandemic. *The Lancet Respiratory Medicine, 8*, 954–955.

The New York Times. (2020). *Covid in the U.S.: Latest map and case count - The New York Times*. Available at: The New York Times https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html. (Accessed 14 November 2020).

United States Census Bureau. (2019). *American community survey 2014-2018 5-year estimates*. Available at: United States Census Bureau https://www.census.gov/newsroom/press-releases/2019/acs-5-year.html. (Accessed 19 May 2020).

Veraset. Analyze + Predict + Build with Veraset's leading global population movement data. Available at https://www.veraset.com/. (Accessed 17 November 2020).

Vermund, S. H., & Pitzer, V. E. (2020). Asymptomatic transmission and the infection fatality risk for COVID-19: Implications for school reopening. *Clinical Infectious Diseases*. https://doi.org/10.1093/cid/ciaa855

Wells, C. R. (2020). Impact of international travel and border control measures on the global spread of the novel 2019 coronavirus outbreak. *Proceedings of the National Academy of Sciences, 117*.

Worby, C. J., & Chang, H.-H. (2020). Face mask use in the general population and optimal resource allocation during the COVID-19 pandemic. *Nature Communications, 11*, 4049.