SIRNET: Understanding Social Distancing Measures with Hybrid Neural Network Model for COVID-19 Infectious Spread

Preprint – Work In Progress

Nicholas Soures[§], David Chambers [‡], Zachariah Carmichael[§], Anurag Daram[§], Dimpy P. Shah[⅄], Kal Clark[⅄], Lloyd Potter[§], Dhireesha Kudithipudi[§]

[§]University of Texas, San Antonio

[‡]Southwest Research Institute

[⅄]University of Texas Health Science Center, San Antonio

[§] [‡] [⅄] MATRIX-AI Consortium

Abstract: The SARS-CoV-2 infectious outbreak has rapidly spread across the globe and precipitated varying policies to effectuate physical distancing to ameliorate its impact. In this study, we propose a new hybrid machine learning model, SIRNET, for forecasting the spread of the COVID-19 pandemic that couples with the epidemiological models. We use categorized spatiotemporally explicit cellphone mobility data as surrogate markers for physical distancing, along with population weighted density and other local data points.

We demonstrate at varying geographical granularity that the spectrum of physical distancing options currently being discussed among policy leaders have epidemiologically significant differences in consequences, ranging from viral extinction to near complete population prevalence. The current mobility inflection points vary across geographical regions. Experimental results from SIRNET establish preliminary bounds on such localized mobility that asymptotically induce containment. The model can support in studying non-pharmacological interventions and approaches that minimize societal collateral damage and control mechanisms for an extended

period of time.

Keywords: SARS-CoV-2, LSTM, SEIR, epidemic models, AI, COVID-19, novel coronavirus, social distancing

1 Introduction

Machine learning techniques have offered solutions to many modeling problems, assuming there is abundant data to train a system [1]. With the rapid impact of COVID-19, several research groups have begun exploring statistical and mathematical models to study the spread of the disease. One of the early studies using AI [2] identified the global spread of the disease through commercial airlines, if the outbreak continues. There are several inefficiencies with the current data available for COVID-19 research, such as limited testing capabilities and high variability within the testing rate (e.g.: 22.08/1000 in Italy, 11.16/1000 in US, to 0.27/1000 in India [3, 4]), inconsistencies in reporting (under reporting), and publicly available data on infection rates currently are unreliable. Particularly lacking is an understanding of the underlying factors which impact the spread, accuracy and availability of reported cases on a small scale, and quantifiable metrics for how social distancing and quarantine efforts impact the spread. To overcome these challenges in providing early models for forecasting the spread of COVID-19, we combine compartmentalized models with a data-driven machine learning approach. In doing so, we address a potential pitfall of machine learning (ensuring compliance with the laws of epidemic dynamics) and a limitation of epidemiological models (enabling the creation of complex mappings from available data sources to critical modeling parameters).

In infectious diseases which are human-human transmissible, individual contact rates drive the spread of infectious pathogens across a population over time. To reduce the high contact rate of COVID-19, several efforts have been taken including social distancing and lockdowns. To account for social distancing and lockdown efforts, we use categorized cell phone mobility

data from Google LLC [5] which includes tracking activity at grocery/pharmacy stores, parks, retail/recreation, residential, train stations, and workplaces globally. The tracking activity is provided in an aggregated format at the country level, state level, and in the US at the county level.

This allows us to implement the epidemic model (which can easily be updated to any of the modules), with a dynamic set of parameters which evolve temporally based on social distancing regulations and other non-pharmaceutical interventions to reduce the spread of COVID-19. This is useful as it allows us to model the role of mobility in managing the spread of COVID-19 and provide insight into potential scenarios if restrictions are lifted by different degrees at certain points in time.

2 SARS-CoV-2

In December 2019, an atypical case of pneumonia was diagnosed in Wuhan [6–8], Hubei province of China which was named COVID-19 and the virus was termed as SARS-CoV-24 [6]. It is a beta coronavirus which is a single-stranded positive sense RNA virus associated SARS-CoV [9, 10]. Because the disease is relatively new, the epidemiology of the virus has changed significantly so far [11–15]. Most cases occur in adults [16]. It is transmitted from person to person through respiratory droplets (within 6 feet) or less likely through contact with fomites [17–19]. The average incubation period is 5.2 days, and each case spreads the infection to an estimated 2.2 people [20]. The median duration of viral shedding has been reported by one study as 20 days with range from 8 to 37 days [21].

First case of COVID19 in US was diagnosed in state of Washington in January 2020 and now U.S. has the highest toll of infected cases and deaths in the world. With no end in sight for the COVID-19 pandemic, it may continue to spread for next several months leading to millions of infections and thousands of deaths before we reach herd immunity. Herd immunity

means when at least 60% of the population is immune from the infection, either through natural exposure to the virus or vaccine administration. Unfortunately, the current infection/exposure estimates for COVID19 are around 1% to 2%.

In the meantime, protective measures such as social distancing, hand washing, avoiding close contact with a sick person, closure of non-essential services, stay at home orders, and covering mouth and nose when around other people are recommended to "flatten the curve" of the epidemic. Without such measures, there can be a high burden on healthcare systems due to the need for hospitalization, intensive care unit admission, and mechanical ventilation. These mitigation efforts, however, do not come without economic and other human consequences. Approximately 10 million Americans applied for unemployment benefits in the last 2 weeks of March and world stock markets lost approximately one-third of their value [22].

3 Importance of Mobility in Modeling

Appropriately tuning mitigation efforts to optimize social welfare has been confounded by the heterogeneous and wave-like pattern of peak-impact globally. The United States, similarly, can expect a high degree of variable impact across metropolitan and micropolitan areas in terms of adherence to policy restrictions on mobility [23], case fatality rates [24], seroprevalence [25], and economic impact [26]. Ultimately policy leaders must, by both necessity and by national policy prescription, translate this variability into an effective reproductive number in order to make optimal policy. The primary challenge posed include the ability to accurately make prescriptive policy in a highly dynamic and geopolitical heterogeneous setting. The capacity to collect, interpret, and model data to make estimates of current and future effective reproductive numbers is now of paramount concern as the United States undertakes phasic pandemic risk mitigation adjustments in order to optimize population morbidity with economic productivity. The stakes cannot be higher: where adjustments are made to decrease overall restrictions on

mobility, assuming all other factors held constant, it will be met with increases in infection. Several recent studies highlight this behavior [27–29]. If the increase in infection is too high, the effective reproductive number will set off a second wave equal to or greater than the initial wave. Similarly, delayed rescindment of restrictions on mobility results in a different type of a second wave: economic disparity, morbidity from inadequately addressed chronic disease, and delayed costs from things like deferred youth education.

Given the above observations, we believe mobility data offers a unique and powerful opportunity to build a valuable model that can answer these challenges. While mobility is most obviously associated with contact rate we believe it has unrealized value via insight into important public health population characteristics which are also dynamic. In particular, mobility data may allow us to indirectly incorporate population features such as health literacy, virtual social substitution (the ability of a population to substitute virtual contact for real contact), and business buy-in [23]. Each of these features may be difficult to measure directly but may be indirectly represented in mobility data.

The mobility data metrics incorporated into this model succeed on four key population modeling criteria which directly (and indirectly) impact the ability to model effective reproductive number: temporally and spatially explicit coverage, representativeness, and contemporaneousness [30]. These metrics are collected in a manner common to all regions which a) allows for rapid wide-scale data interpolation and extrapolation and b) accurately reflects the unique geopolitical profile of regions, each with variable laws, customs, socioeconomic profiles, health care resources, and susceptibility rates. Temporally explicit coverage defines the ability of the mobility data to reasonably describe the movement behaviors of people in the three main environmental contexts of a population *(model representation in parentheses)*: work (workplace), non-work non-residential (retail and recreation, grocery and pharmacy, parks, transit stations), and residential (residential). Spatial specificity refers to the ability to collect granular geograph-

ical data. Representative refers to the ability for the data to offer a high enough sample size to reasonably reflect the population statistic. Contemporaneousness reflects that population statistics are dynamic; the ability of the model to adjust and incorporate changes in behaviors in near real-time permits the model to be prescriptive and responsive to data aggregated across regions. Integration of these metrics into a usable public health resource necessitates a novel and robust modeling strategy.

4 Proposed Hybrid Model: SIRNET

In this research, we focus on learning and forecasting the trends in time series via a hybrid model of neural networks and epidemiological models. The forecasting network, referred to as SIRNET (named after the foundational epidemiological model), learns from i) a sequence of prior trends that carry long-term contextual information (global time-series) and ii) more recent data inputs that are raw (local time-series) and can inform the forecasting of any abrupt changes.

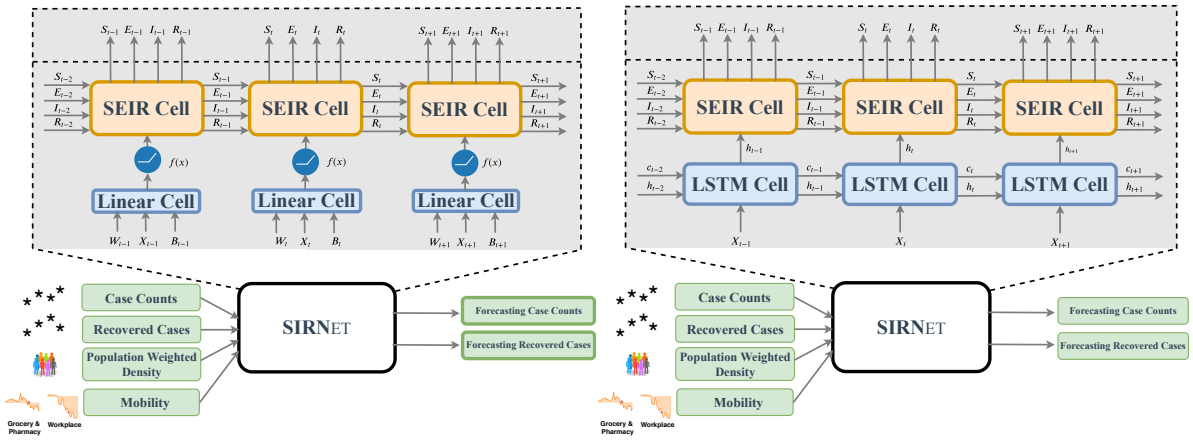

Figure 1: High level visualization of the SIRNET architecture. (a) The RNN is a linear network with input layer $\in \mathbb{R}^6$, hidden layer $\in \mathbb{R}^4$, and output layer $\in \mathbb{R}^1$, with ReLU activation. $\mathbf{B_i}$ is the intractable contact rate. The RNN is a deep LSTM whose internal state is fed as input to the SEIR cell.

4.1 Problem Formulation

Here, we formalize the general form of this forecasting problem. We are given a set of time series that temporally enumerates active, recovered, and fatal cases of COVID-19. The data exhibits varying levels of geographical granularity, i.e., grouping by country, region, sub-region, etc., and irregular onset of the first case. Each sequence $\mathbf{Y} \in \mathbb{R}^{T\times 2}$ comprises T timesteps, which varies between samples, each with the count of active cases and recovered cases. Provided this data, we desire to learn a model that is able to forecast future values of $\mathbf{Y}$. To aid in improving the fit, additional factors should be considered that may impact the infection, recovery, or mortality rates pertaining to the disease. We refer to these additional attributes as features which can be either scalar $\mathbf{x} \in \mathbb{R}^{F}$, spatial $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{F_1\times F_2\times\cdots\times F_n}$, and/or temporal $\mathbf{X} \in \mathbb{R}^{T\times F}$ ($\boldsymbol{\mathcal{X}} \in \mathbb{R}^{T\times F_1\times\cdots\times F_n}$ if spatiotemporal)[1]. With these definitions, the learning problem can be posed as follows (also see (1)). Given historical case data $\mathbf{Y}$ and relevant attributes $\boldsymbol{\mathcal{X}}$ for an area or multiple areas, can we model ($\mathcal{M}$) the latent trends of this data to forecast future cases of COVID-19?

$$\text{minimize} \quad \text{cost}\left(Y_{t+k}, \mathcal{M}(\boldsymbol{\mathcal{X}}_t, Y_t; \theta_0, \theta_1, \dots)\right) \tag{1}$$

4.2 Linear and LSTM Cells

A high-level architecture visualization is shown in Figure 1. The historical time series and local raw data input can be fed to different types of Recurrent neural networks (RNNs) or a Linear cell. Linear cell can be a simple 3-layer network or a deep network with different activation functions. A rectified linear unit works best in most cases. RNNs have the ability to learn patterns from arbitrarily long spatiotemporal data, through cyclic connection of nodes in the network [31]. An LSTM cell, a type of RNN, addresses some of the shortcomings in vanilla RNNs, such as vanishing or exploding parameters by adding memory [32]. The LSTM is

[1] A tensor $\boldsymbol{\mathcal{T}}$ is a multilinear data structure (a higher-order matrix in layman's terms) and is denoted by bold uppercase calligraphic font.

computed by (2) where $[W_{ij}^{hI}, W_{ij}^{I}, W_{ij}^{hf}, W_{ij}^{f}, W_{ij}^{ho}, W_{ij}^{o}, W_{ij}^{hc}, W_{ij}^{c}]$ are a set of weight tensors, h_j^{t-1} vector represents the output of the hidden layer in the previous time step, X_i is the input vector, $[B_j^h, B_j^f, B_j^o, B_j^c]$ represent bias vectors, and θ and Γ are activation functions.

$$\begin{cases} I_j^t = \theta(B_j^h + \sum_{i=0}^{N_1} h_j^{t-1} \times W_{ij}^{hI} + \sum_{i=0}^{N_2} X_i \times W_{ij}^{I}), \\ f_j^t = \theta(B_j^f + \sum_{i=0}^{N_1} h_j^{t-1} \times W_{ij}^{hf} + \sum_{i=0}^{N_2} X_i \times W_{ij}^{f}), \\ O_j^t = \theta(B_j^o + \sum_{i=0}^{N_1} h_j^{t-1} \times W_{ij}^{ho} + \sum_{i=0}^{N_2} X_i \times W_{ij}^{o}), \\ \tilde{C}_j^t = \Gamma(B_j^c + \sum_{i=0}^{N_1} h_j^{t-1} \times W_{ij}^{hc} + \sum_{i=0}^{N_2} X_i \times W_{ij}^{c}), \\ c_j^t = f_j^t \odot c_j^{t-1} + I_j^t \odot \tilde{C}_j^t, \\ h_j^t = O_j^t \odot \Gamma(c_j^t), \\ y_j^t = \gamma(B_j^y + \sum_{i=0}^{N_3} h_j^t \times W_{ij}^{y}) \end{cases} \tag{2}$$

The SIRNET consists of a recurrent neural network to implement the temporal and population dynamics of an SEIR cell, and its framing as such allows us to introduce complex functions with learnable parameters, enabling mapping from salient input data (such as mobility) to the underlying properties of the epidemiological model.

4.3 SEIR Cell

One standard approach to epidemic modeling is compartmentalized models such as SEIR - with Susceptible S, Exposed E (latent infected, but not yet infectious), Infected I, and Recovered R (no longer infectious, also referred to as removed) states. The rate of change in these parameters is represented by the ordinary differential equations (3)-(6) and parameterized by β (effective contact rate/infectious rate learned from mobility data), σ (the incubation rate), and γ (recovery rate).

$$\frac{dS}{dt} = -\beta SI \tag{3}$$

$$\frac{dE}{dt} = \beta SI - \sigma E \tag{4}$$

$$\frac{dI}{dt} = \sigma E - \gamma I \tag{5}$$

$$\frac{dR}{dt} = \gamma I \tag{6}$$

The basic reproduction number representing the number of secondary infections from a primary individual in a completely susceptible population can be computed by,

$$R_0 = \frac{\beta}{\gamma} \tag{7}$$

In the proposed SIRNET model, we use two different neural networks to learn the parameters β and γ, based on population weighted density and cell-phone mobility data (latent information of the contact rate). The mobility data at country, state, and county level is used to predict the contact rate within those regions respectively. In the model, recovery rate can be treated as an individual trainable parameter, or it can be treated as a constant established by medical reporting.

In particular, the model attempts to learn $\beta(t)$ by mapping $\beta(x(t))$, where x represents relevant temporal data (we consider only time steps of one day). The SEIR cell's hidden states consist of the four compartmental groups normalized by population.

While our approach can be extended to many types of data, our work here is focused on one type in particular: mobility data. Contact rate is a key parameter of the model and its modification through quarantine measures is an effective way to control the spread of the virus. Contact rate is a function of population density as well as how people move and interact with each other. Traditional modeling can retrospectively estimate the change in contact rate brought about by policy changes (step-function changes), in our approach we build upon this technique to allow the integration of richer, daily information based on the actual activities of a population. To this end, we begin with cell-phone based mobility information.

The mobility input vector, x, consists of mobility ratios (current mobility divided by nominal mobility) in 6 categories provided through [5]. SIRNET's task is to use this feature vector to learn the resulting contact rate as a function of population mobility. Through the use of the

SEIR cell, we can map the output to case counts and learn the underlying mobility to contact rate function in an end-to-end fashion.

For the mobility model, the SEIR cell predicts contact rate according the following function:

$$\beta(t) = \text{ReLU}(W \cdot x)^p \tag{8}$$

Our model is a linear combination of mobility contributions to an effective mobility, raised to the power p. The parameterization of p allows the model to learn the effective power of scaling mobility rather than simply assuming one, removing the need to justify a linear or quadratic relationship (our modeling exercises suggest the latter provides the best representation). The rectified linear unit function ensures that our model will only produce non-negative contact rates.

One of the primary challenges in modeling is that the underlying state is difficult to estimate. The only data that is reliably available is the total case count, and the case count only represents a fraction of actual cases (with an estimated 50%-80% being underreported). It also lags the true state of the system by several days. Mobility data will drive exposure, exposure will drive the amount of the infectious, and the infections will in turn drive the number of cases. To account for all of these factors, we use the 5-day incubation period [33–35] and add an additional 5 days to account for the delay between becoming infectious and receiving a positive test confirmation. This delay in testing is not constant across time, nor is it consistent from location to location, but the measurable impacts of mobility on contact rate are most apparent when delay is taken into account.

We initialize the hidden state with the number of active cases at the onset of the epidemic, I_0.

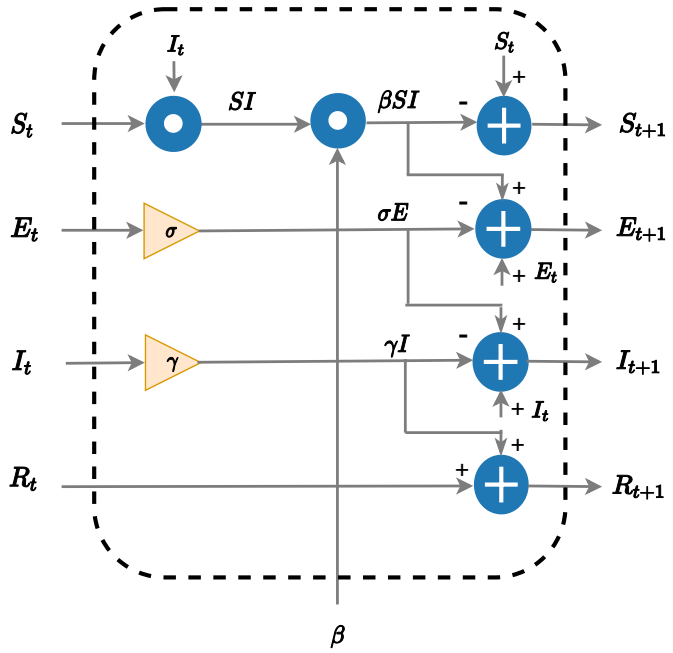

Figure 2: We frame the SEIR modeling as a recurrent neural network (RNN) architecture, introducing the SEIR-cell which encodes susceptible, exposed, infected, and recovered proportions in its hidden states. Open sourced on *GitHub* – download here.

4.4 Assumptions

The model uses case count resources from WHO [36], CDC [37], European CDC [38], NYTimes [39], and Texas DSHS [40]. Specific data reported includes confirmed cases, deaths, and recovered cases. The population weighted density data and age-group data for US is obtained from the census website [41] and mobility data is captured from here [5] and here [42]. The assumptions made in the model are listed in Table 1. Additionally, the effective social distancing measures for each of the regions are populated from here [43]. Testing and recovery data is collected from multiple sources for cross-validation, such as [43–45].

Table 1: Parameters used in the models

Parameter	Real-world Value	Experimental	References
Incubation Period	≈ 5.2 days	-	[33–35]
Viral Shedding	≈ 8-37 days		[21]
Acute hospital beds	-	40% of total	[46]
Testing Delays	≈ 10 days		[4]
Initial infected	First reported case	-	

5 Results and Analysis

Model analysis for different geographical regions:

The SIRNET was evaluated on different geographical regions. Figure 3 shows the fit for predicting total cases in a region based on mobility data compared to the ground truth in several countries. Figures 4 and 5 show the models forecast for different mobility levels in each country for active cases and total cases respectively. These figures demonstrate that the proposed SIRNET is able to fit the case count by region well using mobility information to determine the contact rate of an SEIR cell. Based on the projected forecasts, we observe that a continuation of quarantine level mobility will result low case counts. If the mobility restrictions are reduced to 50% nominal mobility, the model shows that this is near the edge of stable peak cases where in some scenarios the curve stays at a low peak, while for others the peak increases drastically compared to the quarantined mobility and occurs much later. The third scenario is 75% nominal mobility which based on the model is expected to result in a slightly delayed peak approximately 2/3 the maximum peak during normal mobility. The exception is South Korea, where at this stage even a return to 75% mobility is not expected to result in a second wave. A

zoomed in figure of the forecast of active cases in the US is shown in Figure 6 to better illustrate the difference between mobility levels. In general, these results suggest a continuation of quarantine level mobility or at least below 50% nominal mobility for the immediate future. Figure 7 shows the model fit with ground truth for the top 28 counties in Texas. In general the model fits well where the case counts are higher than 50 and as data becomes richer, the fit improves significantly. Figure 8 highlights a sample of the mobility trends used in all our simulations. **It is important to note that this data reflects a sample space of mobility for the region and might be missing information on key populations that do not use specific types of devices. Adding finer granular information and data from multiple data providers can alleviate this concern, as it becomes open-sourced.**

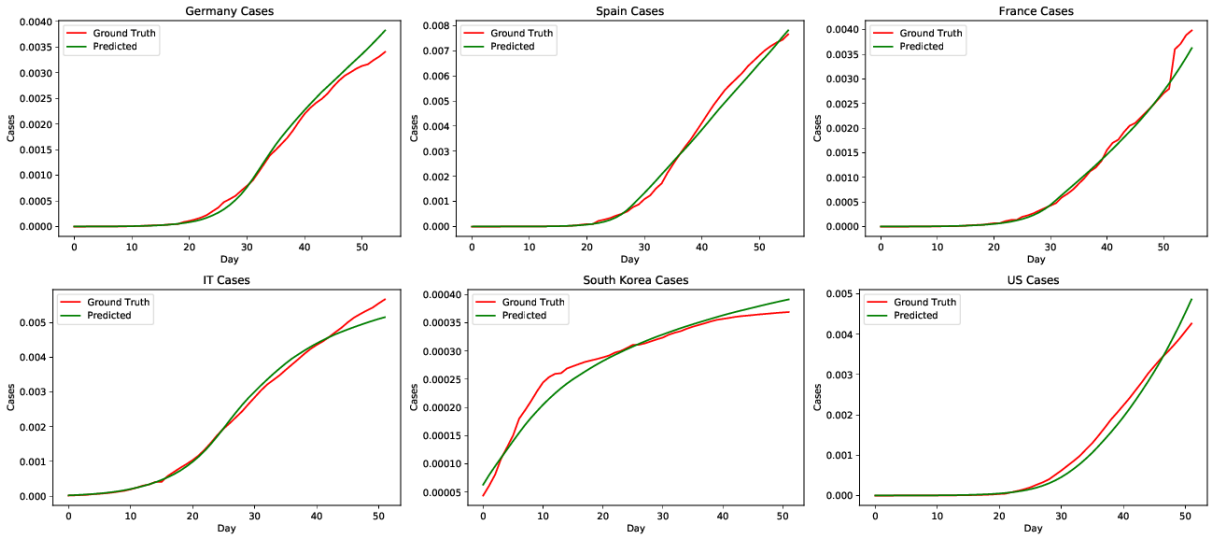

Figure 3: SIRNET predictions on previous and current CDC reported cases data for the United States, Italy, Spain, Germany, France, and South Korea.

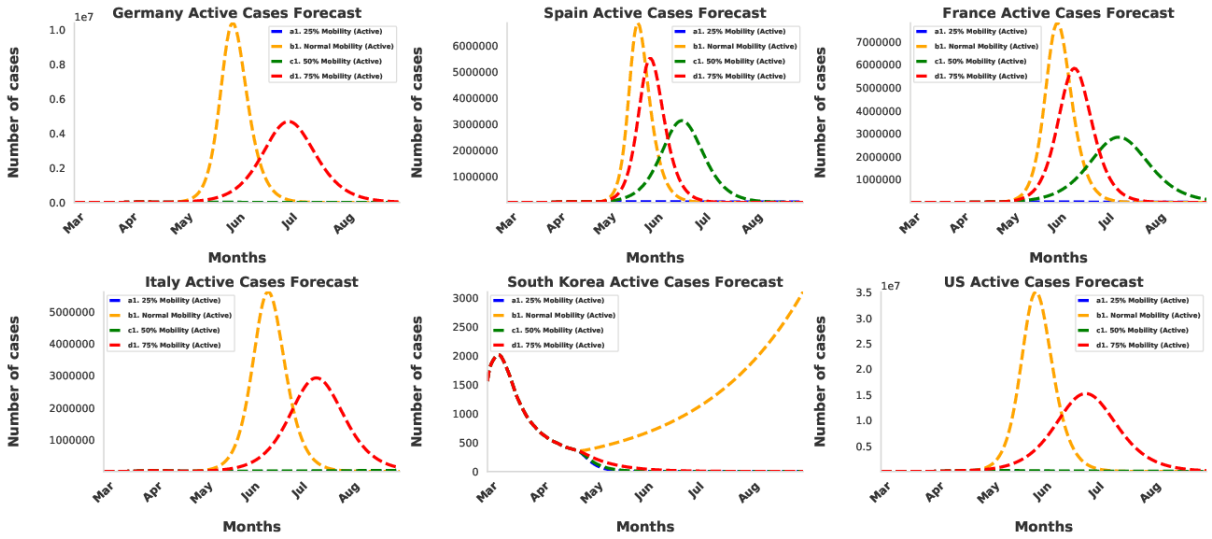

Figure 4: SIRNET forecast of active cases for the United States, Italy, Spain, Germany, France, and South Korea.

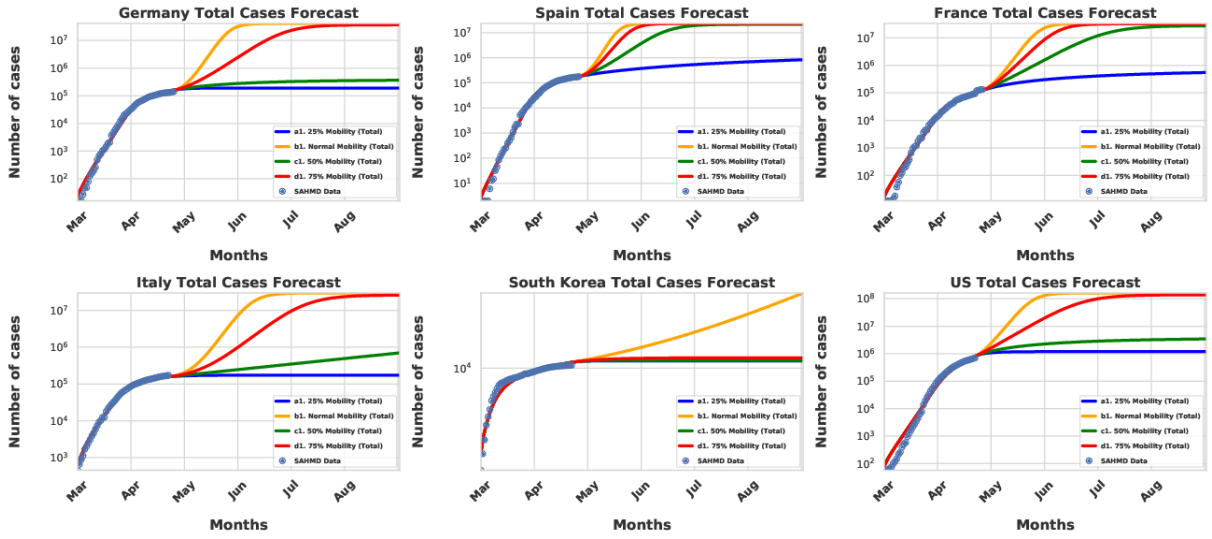

Figure 5: SIRNET forecast of total cases for the United States, Italy, Spain, Germany, France, and South Korea.

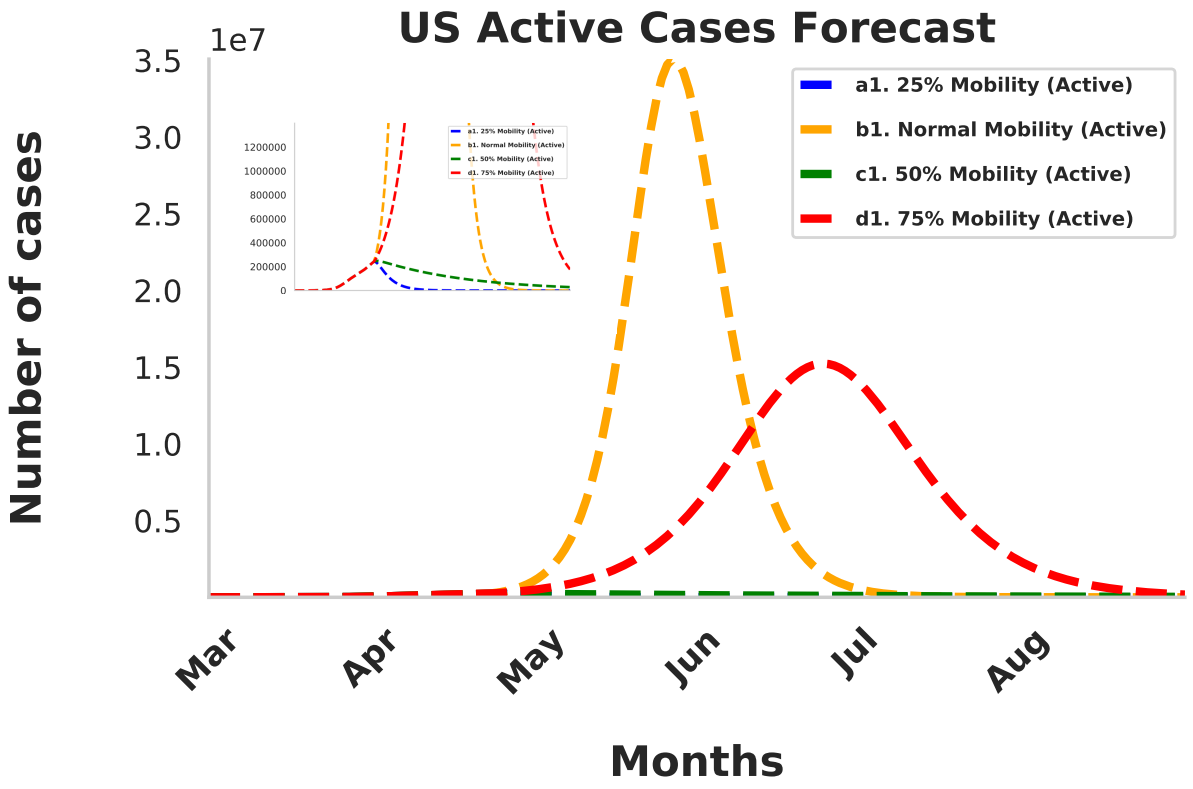

Figure 6: Forecasting of active cases for USA under four different scenarios of physical distancing. Mobility data is real-time cell phone/mobile device location for USA collected from here [5]. Inset on the top left is further magnified onto the active cases with mobility of 25% and 50%. The scale is reflective of this change.

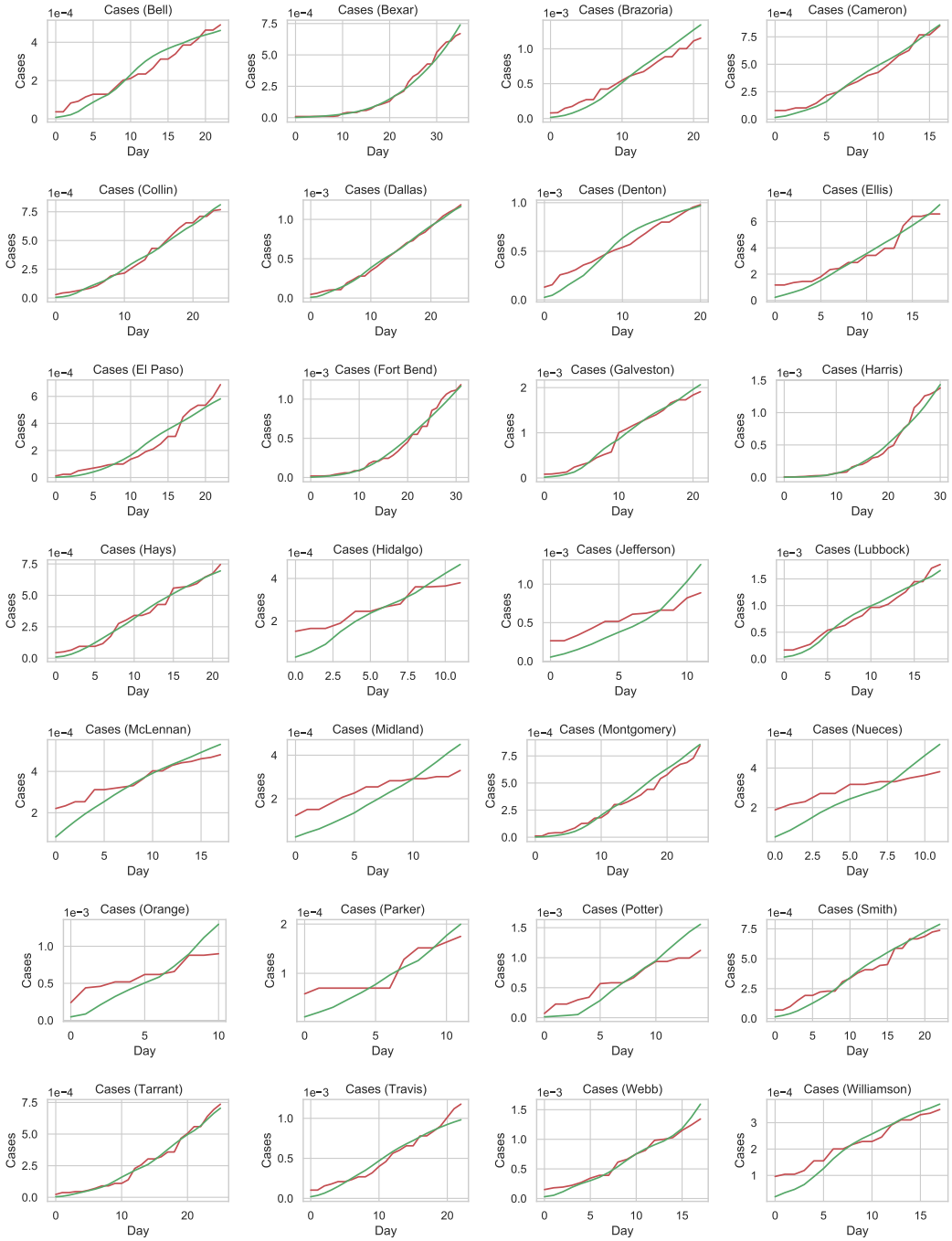

Figure 7: Model fit across the top 28 counties in TX, with all six mobility data points available. The physical distancing scenario is trained on 51 regions around the world encapsulating different levels of intervention. The red line is the ground truth total case data and the green line is the predicted total case data.

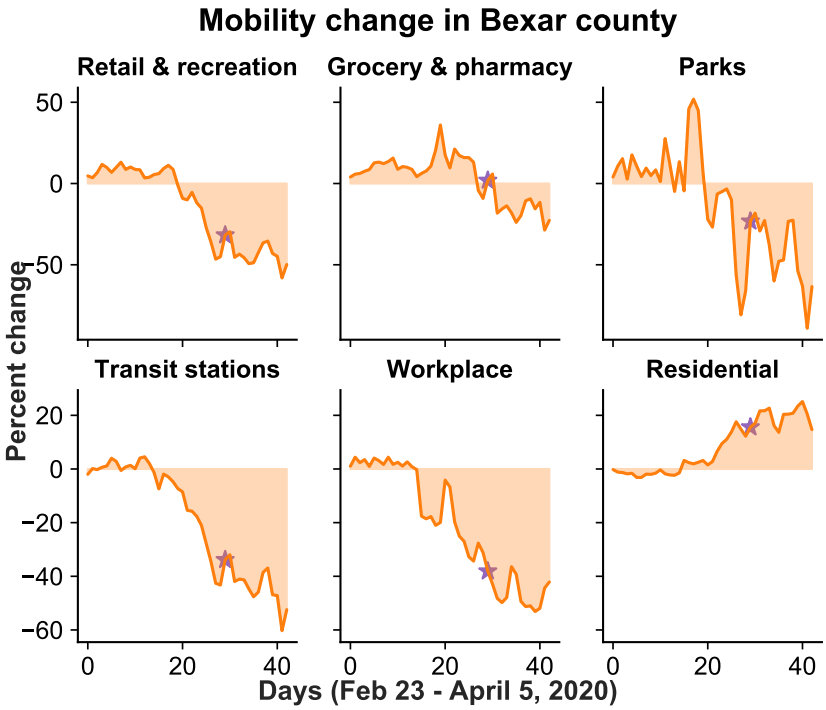

Figure 8: Sample mobility data changes in Bexar County, Texas from Feb 23rd to April 11th. The star indicates when social distancing measures were implemented in the county (March 23rd). Data is collected using webscraping tools developed by the team. By comparison, mobility in Italy has decreased by ≈ 70% to non-essential locations, with the lockdown procedures. Data Source: Google LLC [40]

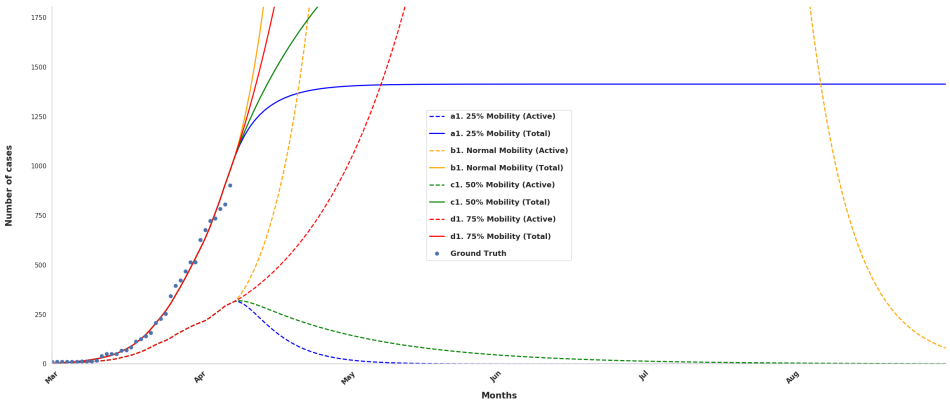

Figure 9: Forecast for Bexar County, TX. All models are driven initially by recorded mobility and then split to forecast different scenarios. While current mobility (as of 11 April 2020) is approximately 50% of nominal, relaxing measures at this time could result in runaway growth.

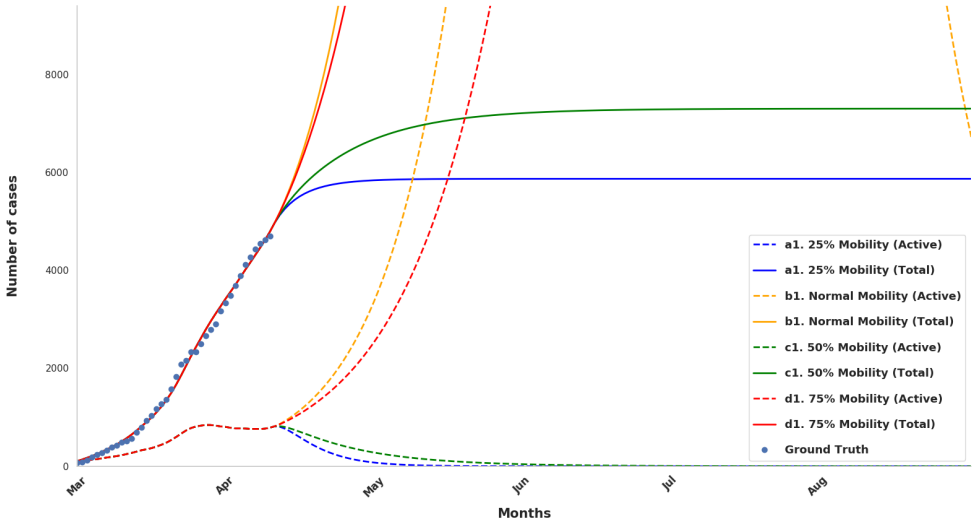

Figure 10: Forecast for King County, WA, where a longer period of infection enables better model fitting. King County's heavy mobility restrictions are clearly dropping the R_0 below 1.

Why Does Reduced Mobility Help?

Our model shows that the epidemic is highly sensitive to changes in mobility rates. Figure 9 and Figure 10 shows examples of two counties at different peak points of spread and how changing mobility rate is impacting these counties (Bexar county in Texas and King County in WA). SEIR cell dynamics show the critical points in reproduction number, in which we project mobility levels showing three possible outcomes. First, heavy mobility restrictions can drive the contact rate and reproduction number down so that the virus is eliminated from the population. The number of active cases and the number of total cases are kept to minimal levels, though the population remains susceptible to second waves. Second, at higher levels of mobility (those approaching nominal), the virus reproduces prolifically in the population, with exponential growth ending as herd immunity is reached. This results in both overwhelmed medical resources due to active cases and a large number of deaths due to the total case count. A third possibility is a reproduction number of approximately 1. In this scenario, active case count

remains low; however, the virus would continue to work its way through the population until herd immunity is reached, resulting in a large number of total cases and deaths as a slow tail gradually tapers off. **In the mobility scenarios tested across countries or counties, mobility > .7 leads to an uncontained outbreak, mobility < .5 results in a local elimination of the virus, and those in between having slower peaks.**

Remobilizing a population will require careful monitoring to ensure that the critical reproduction number is not approached. It should also be considered that while our model introduces a new feature for relating human activity to mobility, it is certainly not the only factor to consider. Changes in behavior, including social distancing and hygiene habits, are certainly also contributing to the reduced reproduction rate in a way that cannot be considered independently (their effects are reflected in mobility data, but are difficult to predict in future mobilization). It is imperative that policymakers consider the dynamic nature of infection. **The current projected peak is the worst time to begin relaxing quarantine measures. Also, there is only one peak for current mobility because of the drastic measures that have been taken thus far**. Continuous monitoring and modeling will be indispensable tools for containing the outbreak, and models will continue to improve as more data becomes available.

Based on the projections for a region, the SIRNET model can calculate the potential hospitalization rates by age group, for best and worst case scenarios. This is computed based on local age demographics, forecasted active cases, and estimated hospitalization rates by age group from experimental data. An example scenario for a county is shown in Figure 11.

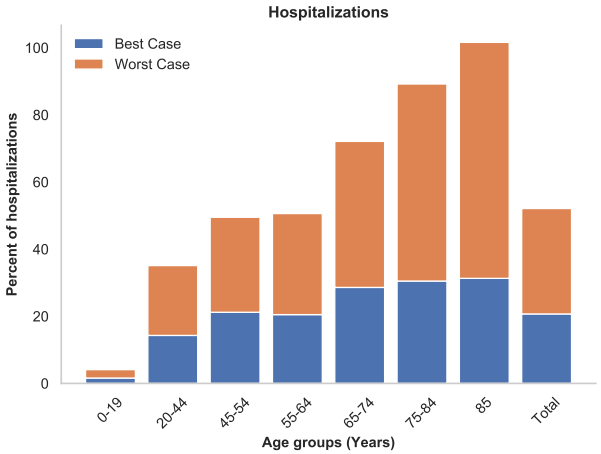

Figure 11: Hospitalization rate for SARS-CoV-2 in a sample county (Bexar, TX) based on age group. The population by age group data for 2018 is obtained from here [41].

Error Tolerance

When using any ML or statistical models to forecast trends, it is important to consider the confidence interval or margin of error for the predictions. SIRNET is currently trained on a specific region it is forecasting, with region-specific assumptions about under-reporting, delay in reporting, the recovery rate, and the transition rate from exposed to infected. In future work it will be necessary to account for the error range for each of these variables based on global reported data, and use this to predict the potential fluctuation in forecasted scenarios. Another important extension to SIRNET is to extend learning to multiple regions, providing a more generalized forecast that can capture distinctions between different regions.

Conclusions and Discussion

Our work takes a multidisciplinary approach to address modeling the spread of COVID-19. SIRNET is a hybrid between epidemic modeling, physical science, and machine learning. The benefit of epidemic modeling, is constraining our network to produce meaningful variables from a physical standpoint which adds an intuitive understanding of how the model is fore-

casting and provides an approach for overcoming limited or missing real-world data samples. On the other hand, machine learning provides a tool for translating variables, such as mobility, non-pharmaceutical intervention, and population demographics, into variables that impact an epidemic model. It also allows us to discover relationships between real-world trends and the impact on the spread of COVID-19, as well as model scenarios such as relaxing social-distancing policies. We believe both components are necessary to develop an insightful model to aid in understanding the impact of non-pharmaceutical interventions on COVID-19.

Similar to other approaches, we base our study on several biologically observed data and real-world datasets. We demonstrate how new tools can be created to better exploit available quantitative measures in the fight against COVID-19. By integrating reliable metrics and well-studied infection dynamics, we create an approach that is deeply data-driven and science-based. Our studies confirm the effectiveness of reduced mobility for limiting the reach of the pandemic, and our models provide a means of forecasting the effects of different mobility scenarios.

SIRNET is in an early iteration and requires extensive sensitivity analysis to understand the range of impact of different parameters. Additionally, exhaustive mobility data combined with non-pharmacological intervention datasets can improve the network predictions. Since several datasets are proprietary and limited by data user agreements, it will be important to establish good data collection and standardization practices to address catastrophic events.

Given the substantial risk of reintroduction of the SARS-CoV-2, it is critical to reinforce balanced social distancing measures in the coming months to reduce the impact on the healthcare system, general public, and economic prosperity. Resource limitations in a rapidly growing pandemic demand compelling resource utilization choices. Of importance is to note that the data-driven AI models provide a window into understanding the potential impact and should be treated as a qualitative guidance due to the rapid changes associated with the data collection, testing strategies, reporting, and the virus transmission.

Acknowledgements

Authors would like to thank all the open-source data providers, which was critical in timely analysis of the spread. We are grateful for researchers S. Hamed Fatemi Langroudi, Pankil Shah (UTHSCSA), and Tej Pandit who have helped in the simulations and tool flows. The authors would like to acknowledge the research organizations and their leads at UTSA, UTHSCSA, and SwRI for their thought leadership and open support for multidisciplinary collaboration. Research team also recognizes several members of the community in the city of San Antonio and UT System who have aided in data access and critical discussions in understanding the virus. We appreciate members of Project Alpha, Amina Qutub and Hongjie Xie, who continue to share insights on COVID-19 recovery and tracking tools.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Z. Obermeyer, E. J. Emanuel, Predicting the future—big data, machine learning, and clinical medicine, *The New England journal of medicine* **375**, 1216 (2016).

2. Pneumonia of unknown aetiology in wuhan, china: potential for international spread via commercial air travel, `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7107534/`. (Accessed on 04/19/2020).

3. To understand the global pandemic, we need global testing – the our world in data COVID-19 testing dataset - our world in data, `https://ourworldindata.org/{COVID}-testing`. (Accessed on 04/19/2020).

4. Why some COVID-19 tests in the us take more than a week | mit technology review, https://www.technologyreview.com/2020/04/05/998460/{COVID}-19-test-results-faster-commercial-labs-delays-coronavirus/. (Accessed on 04/13/2020).

5. COVID-19 community mobility reports, https://www.google.com/{COVID}19/mobility/. (Accessed on 04/18/2020).

6. A. J. Rodriguez-Morales, *et al.*, Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis, *Travel medicine and infectious disease* p. 101623 (2020).

7. T. Singhal, A review of coronavirus disease-2019 (COVID-19), *The Indian Journal of Pediatrics* pp. 1–6 (2020).

8. R. Ralph, *et al.*, 2019-nCoV (Wuhan virus), a novel coronavirus: Human-to-human transmission, travel-related cases, and vaccine readiness, *The Journal of Infection in Developing Countries* **14**, 3 (2020).

9. A. Gorbalenya, *et al.*, Coronaviridae study group of the international committee on taxonomy of viruses: The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2, *Nature microbiology* **2020**, 03 (2020).

10. X. Zou, *et al.*, Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection, *Frontiers of medicine* pp. 1–8 (2020).

11. P. Sun, X. Lu, C. Xu, W. Sun, B. Pan, Understanding of COVID-19 based on current evidence, *Journal of medical virology* (2020).

12. E. Callaway, D. Cyranoski, China coronavirus: Six questions scientists are asking., *Nature* **577**, 605 (2020).

13. Y.-R. Guo, *et al.*, The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak–an update on the status, *Military Medical Research* **7**, 1 (2020).

14. P. Habibzadeh, E. K. Stoneman, The novel coronavirus: a bird's eye view, *Int J Occup Environ Med (The IJOEM)* **11**, 1921 (2020).

15. S. P. Adhikari, *et al.*, Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review, *Infectious diseases of poverty* **9**, 1 (2020).

16. C. China, Novel coronavirus pneumonia emergency response epidemiology team, *Vital surveillance: The epidemiological characteristics of an outbreak of* (2019).

17. N. van Doremalen, *et al.*, Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1, *New England Journal of Medicine* (2020).

18. H. Aydin, D. Al-Khooly, J. E. Lee, Influence of hydrophobic and electrostatic residues on SARS-coronavirus s2 protein stability: Insights into mechanisms of general viral fusion and inhibitor design, *Protein Science* **23**, 603 (2014).

19. Y. T.-C. Yu, *et al.*, Surface vimentin is critical for the cell entry of SARS-CoV, *Journal of biomedical science* **23**, 14 (2016).

20. Q. Li, *et al.*, Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia, *New England Journal of Medicine* (2020).

21. F. Zhou, *et al.*, Clinical course and risk factors for mortality of adult inpatients with COVID-19 in wuhan, china: a retrospective cohort study, *The Lancet* (2020).

22. A. Fitzgerald, K. Kwiatkowski, V. Singer, S. Smit, An instant economic crisis: How deep and how long? Library Catalog: www.mckinsey.com.

23. R. K. Webster, *et al.*, How to improve adherence with quarantine: Rapid review of the evidence, *Public Health* (2020).

24. D. D. Rajgor, M. H. Lee, S. Archuleta, N. Bagdasarian, S. C. Quek, The many estimates of the COVID-19 case fatality rate, *The Lancet Infectious Diseases* (2020).

25. E. Bendavid, *et al.*, COVID-19 Antibody Seroprevalence in Santa Clara County, California, *medRxiv* p. 2020.04.14.20062463 (2020). Publisher: Cold Spring Harbor Laboratory Press.

26. D. Wheaton, Explore the economic impact of COVID-19 with these charts (2020). Library Catalog: www.sandiegouniontribune.com Section: Business.

27. J. R. Koo, *et al.*, Interventions to mitigate early spread of sars-cov-2 in singapore: a modelling study, *The Lancet Infectious Diseases* (2020).

28. A. J. Kucharski, *et al.*, Early dynamics of transmission and control of covid-19: a mathematical modelling study, *The lancet infectious diseases* (2020).

29. L. Ferretti, *et al.*, Quantifying dynamics of sars-cov-2 transmission suggests that epidemic control and avoidance is feasible through instantaneous digital contact tracing, *medRxiv* (2020).

30. P. Deville, *et al.*, Dynamic population mapping using mobile phone data, *Proceedings of the National Academy of Sciences* **111**, 15888 (2014). Publisher: National Academy of Sciences Section: Physical Sciences.

31. Z. C. Lipton, J. Berkowitz, C. Elkan, A critical review of recurrent neural networks for sequence learning, *arXiv preprint arXiv:1506.00019* (2015).

32. J. Schmidhuber, S. Hochreiter, Long short-term memory, *Neural Computation* **9**, 1735 (1997).

33. The Incubation Period of COVID-19 From Publicly Reported Confirmed Cases | Annals of Internal Medicine | American College of Physicians, https://annals.org/aim/fullarticle/2762808/incubation-period-coronavirus-disease-2019-{COVID}-19-from-publicly-reported. (Accessed on 04/12/2020).

34. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia | NEJM, https://www.nejm.org/doi/full/10.1056/NEJMoa2001316. (Accessed on 04/12/2020).

35. Clinical Characteristics of Coronavirus Disease 2019 in China | NEJM, https://www.nejm.org/doi/full/10.1056/NEJMoa2002032. (Accessed on 04/12/2020).

36. Situation reports, https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports. (Accessed on 03/30/2020).

37. Coronavirus Disease 2019 (COVID-19) | CDC, https://www.cdc.gov/coronavirus/2019-nCoV/index.html. (Accessed on 03/30/2020).

38. COVID-19, https://qap.ecdc.europa.eu/public/extensions/{COVID}-19/{COVID}-19.html. (Accessed on 03/30/2020).

39. NYtimes/COVID-19-data: An ongoing repository of data on coronavirus cases and deaths in the U.S., https://github.com/nytimes/{COVID}-19-data. (Accessed on 03/30/2020).

40. COVID-19 In Texas (Dashboard), https://txdshs.maps.arcgis.com/apps/opsdashboard/index.html#/ed483ecd702b4298ab01e8b9cafc8b83. (Accessed on 03/30/2020).

41. Population, https://www.census.gov/topics/population.html. (Accessed on 04/21/2020).

42. Covid-19 - mobility trends reports - apple, https://www.apple.com/covid19/mobility. (Accessed on 04/21/2020).

43. Covid-19, https://covid19.healthdata.org/united-states-of-america. (Accessed on 04/21/2020).

44. Coronavirus disease (covid-19) - statistics and research - our world in data, https://ourworldindata.org/coronavirus. (Accessed on 04/21/2020).

45. Csse – center for systems science and engineering at jhu, https://systems.jhu.edu/. (Accessed on 04/21/2020).

46. COVID act now, https://{COVID}actnow.org/. (Accessed on 04/13/2020).

47. Stay-at-home orders to fight COVID-19 in the united states: The risks of a scattershot approach | the henry j. kaiser family foundation, https://www.kff.org/coronavirus-policy-watch/stay-at-home-orders-to-fight-{COVID}19/. (Accessed on 04/13/2020).

Supplementary materials

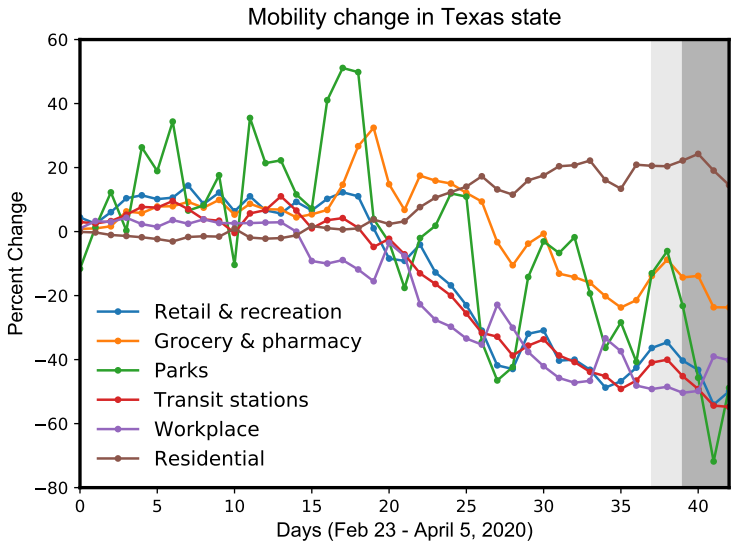

Figure 12: Mobility changes in the state of Texas over the past 6 weeks. The most recent datapoint is April 5th. Data is collected using webscraping tools developed by the team. Mobility to workplace and residential areas has decreased ~42%, over the past 40 days. The shaded regions represent the days when stay at home order was announced (March 31st) and was effective (April 2nd) in Texas [47]. Data Source: Google LLC [5]