COVID-19 Risk Estimation using a Time-varying SIR-model

Mehrdad Kiamari
Viterbi School of Engineering,
University of Southern California
Los Angeles, USA
kiamari@usc.edu

Gowri Ramachandran
Viterbi School of Engineering,
University of Southern California
Los Angeles, USA
gsramach@usc.edu

Quynh Nguyen
Viterbi School of Engineering,
University of Southern California
Los Angeles, USA
quynhu@usc.edu

Eva Pereira
Office of the Mayor,
City of Los Angeles
Los Angeles, USA
eva.pereira@lacity.org

Jeanne Holm
Office of the Mayor,
City of Los Angeles
Los Angeles, USA
jeanne.holm@lacity.org

Bhaskar Krishnamachari
Viterbi School of Engineering,
University of Southern California
Los Angeles, USA
bkrishna@usc.edu

ABSTRACT

Policy-makers require data-driven tools to assess the spread of COVID-19 and inform the public of their risk of infection on an ongoing basis. We propose a rigorous hybrid model-and-data-driven approach to risk scoring based on a time-varying SIR epidemic model that ultimately yields a simplified color-coded risk level for each community. The risk score Γ_t that we propose is proportional to the probability of someone currently healthy getting infected in the next 24 hours based on their locality. We show how this risk score can be estimated using another useful metric of infection spread, R_t, the time-varying average reproduction number which indicates the average number of individuals an infected person would infect in turn. The proposed approach also allows for quantification of uncertainty in the estimates of R_t and Γ_t in the form of confidence intervals. Code and data from our effort have been open-sourced and are being applied to assess and communicate the risk of infection in the City and County of Los Angeles.

CCS CONCEPTS

• **Computing methodologies → Modeling methodologies**.

KEYWORDS

Risk Modelling, COVID-19, SIR model

ACM Reference Format:
Mehrdad Kiamari, Gowri Ramachandran, Quynh Nguyen, Eva Pereira, Jeanne Holm, and Bhaskar Krishnamachari. 2020. COVID-19 Risk Estimation using a Time-varying SIR-model. In *1st ACM SIGSPATIAL International Workshop on Modeling and Understanding the Spread of COVID-19 (COVID-19), November 3, 2020, Seattle, WA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3423459.3430759

ACM acknowledges that this contribution was authored or co-authored by an employee, contractor, or affiliate of the United States government. As such, the United States government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for government purposes only.
COVID-19, November 3, 2020, Seattle, WA, USA
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-8168-0/20/11…$15.00
https://doi.org/10.1145/3423459.3430759

1 INTRODUCTION

The ongoing COVID-19 epidemic has forced governments and public authorities to employ stringent measures [6, 10], including closing business and implementing stay-at-home orders, to contain the spread. When making such decisions, policymakers require tools to understand in "real-time" how the virus is spreading in the community, as well as tools to help communicate the level of risk to citizens so that they can be encouraged to take appropriate measures and take the public health directives seriously.

One metric that has been found to be useful for authorities to assess the level of containment over time is the effective reproduction number [7]. The effective reproduction number, R_t, indicates on average how many currently susceptible persons can be infected by a currently infected individual. The epidemic grows if this measure is above one. It is desirable to keep this value as far below one as possible over time in order to contain and eventually, hopefully, eliminate the virus from the community.

While R_t is meaningful to understand the rate at which the epidemic is spreading and has been proposed previously (for example, see https://rt.live/), what has been missing in the public discourse is a risk metric that is more suitable for communication to a wider public. One key requirement for such a metric is that it be something that a citizen could relate to on an individual basis. Another requirement is that it needs to be easy to communicate to a wide audience. We address both these requirements in this work and make the following contributions.

First, we obtain the daily effective reproduction number R_t of a time-varying SIR model as well as the corresponding confidence Interval. The confidence interval reflects uncertainty in both the parameter of the underlying model and uncertainty in the data itself. Further, we present the mathematical derivation of the distribution of R_t.

Second, we propose a novel risk score Γ_t for a community that is proportional to the probability that an individual will get infected in the next 24 hours. We show that the risk score can be calculated given estimates of four quantities: a) an estimate of $I_{rep,new}(t)$, the most recently reported count of new confirmed infectious cases, b) an estimate of R_t as discussed above, c) an estimate of K, the ratio of true infectious cases to the number of confirmed cases, and d) an estimate of $S(t)$, the current number of susceptible individuals in the community. To make the score more meaningful, we normalize

the probability of infection by multiplying it by 10,000. Then, a risk score of x is an indication that there is, on average, a chance of x in 10,000 of an individual in the community becoming infected in the next 24 hours. We also propose to convert the numerical risk score, which has an intuitive meaning as indicated above, to a color-coded risk level based on suitably chosen thresholds[1]. We propose the use of four color-levels to indicate the corresponding risk level from very low to high: green, yellow, orange, and red.

Third, we have implemented software to estimate the risk level for any community and released it as open-source. The code requires only time-series data on confirmed new cases, the population of the community, and an estimate for the ratio of true to confirmed (detected) COVID-19 positive cases. This software is being used at USC to process the daily data of communities within Los Angeles County to estimate and generate maps of risk levels by community. The block diagram in figure 1 illustrates key elements of our system design. Our data parser is able to get the raw data from online data sources, clean them up and store them in machine-friendly (csv and json) formats. Our code for infection risk calculation uses this data in conjunction with a time-varying SIR-based Bayesian mathematical model to obtain risk estimates and prediction for different communities. The results are provided in CSV format and can be used to generate a heatmap-type visualization as well.

The risk scoring model we describe in this work is now being used by the City of Los Angeles, which in turn is working with the County of Los Angeles and other partners to develop a publicly accessible tool that can be used by individuals and communities to grow awareness and mitigate risk of infection. We believe that our risk estimation approach will be similarly of value to other communities around the world.

The rest of the paper is structured as follows: Section 2 reviews the related work. The novel risk calculation methodology is presented in Section 3. Section 4 discusses the implementation and evaluation of the proposed method in Los Angeles County. The key results are discussed in Section 5. Section 6 concludes the paper.

2 RELATED WORK

There have been a few recent works studying different transmission models for COVID-19 such [4] which developed an agent-based model to reproduce the characteristics of COVID-19 transmission or [1] which proposed a mobility-based model to measure COVID-19 growth rate ratio for a given day.

As noted above, the calculation of the risk score requires an estimate of R_t. We show how this can be estimated using a time-varying SIR model, a generalization of the well-known SIR compartmental model [3, 8] which consists of three states, namely the susceptible state, the infected state, and the recovered state. While traditionally this model is assumed to have a interaction rate / infection rate parameter that is constant, one recent work has used a time-varying SIR model to recover the time-varying effective reproduction number [5]. Going beyond that work, we also show how to derive a confidence interval for R_t in this work. Further, the authors of [5] make strong assumptions on the number of susceptible individuals by approximating it as a constant factor of the entire population. This assumption may not be accurate when the number of infected individuals are high compared to the total population of a community; we therefore take a more general approach.

Another recent work by Systrom [9] has presented a Bayesian prediction approach to obtain confidence intervals for R_t. However, Systrom's work builds on [2], where the definition of infection rate R_t is not based on a time-varying contact rate of the SIR model. Instead, their approach estimates infection rate probabilistically based on the number of new cases alone.

We are not aware of prior work that has proposed defining risk for COVID-19 or other epidemics in terms of an individual's probability of infection, which we argue is more meaningful for communicating risk to the public.

3 METHODOLOGY

Compartmental mathematical models for epidemic spreads including the well-known SIR model have been used since the work of Kermack and McKendrick in 1927 [3]. In the SIR model, each member of a given population is in one of three states at any time: susceptible, infectious, recovered. Any individual that is susceptible could become infected with some probability when they come into contact with an infected individual. Any individual that is infectious eventually recovers (in the context of COVID-19 when applying the SIR model, note that the category of recovered individuals will also include removed individuals due to deaths, which could be modeled as a constant fraction of all individuals in this category). In the classical SIR model, the number of susceptible individuals that become infected depends on the rate at which infected and susceptible individuals encounter each other and this rate is assumed to be constant. A well-known parameter in the classical SIR model is called R0, the effective reproductive number, which measures the average number of infections caused by infectious individuals at the beginning of the epidemic.

Time-Varying SIR model and R_t: In our work, we have extended the SIR model to a time-varying model, in which the rate of encounters and infection probability between individuals in the population is assumed to be time-varying. This better reflects the reality of our present epidemic where interventions such as stay-at-home have been put in place and relaxed and various times and compliance with recommendations such as wearing masks and maintaining physical density has also been time-varying. Based on this model, we are able to define and derive a new approach to calculating a time-varying version of the effective reproductive number, which we refer to as R_t.

A particularly innovative aspect of our model is that it is a Bayesian model that allows the incorporation of various sources of uncertainty in the model, including uncertainty in the actual numbers of infected individuals (due to not every infected individual having been tested, as studies [2] have shown), uncertainty in recovery times, and uncertainty in the choice of parameters for de-noising the empirical data. This allows us to generate not only an estimate of R_t, but also quantify confidence in the estimate from a rigorous statistical perspective.

In this section, we elaborate upon the SIR model in detail. The SIR model is one of the simplest and the most well-known epidemic model [3, 8] where each person belongs to one of the following three

[1]Thresholds for categorizing into very low-risk, low-risk, medium-risk, and high-risk levels are set from a medical perspective.

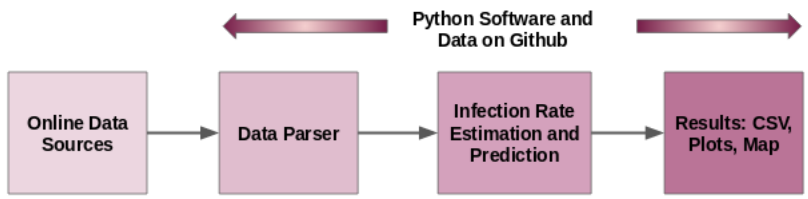

Figure 1: Overview of Our System.

states: the susceptible state, the infected state, and the recovered state. Regarding the susceptible state, individuals have not had the virus yet. However, they may get infected in case of being exposed to an infected individual. As far as the infected state is concerned, a susceptible person has the virus after being exposed to infected individuals. Finally, a person enters the recovered state in case of either the individual gets healed or dies. One important point about this model is that a recovered person will not be a susceptible one anymore. This is how the model is constructed, as in most cases it appears that COVID-19 has an extremely low re-infection rate, at time of writing this paper.

The SIR model follows the following differential equations:

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta \frac{S(t)I(t)}{N} \\ \frac{dI(t)}{dt} &= \beta \frac{S(t)I(t)}{N} - \sigma I(t) \\ \frac{dR(t)}{dt} &= \sigma I(t) \end{aligned} \tag{1}$$

where $S(t)$, $I(t)$, and $R(t)$ respectively represent the number of susceptible, infected, and recovered people in a population size of N at time t. Regarding the parameter σ, it is the recovery rate after being infected and is equal to $\frac{1}{D_I}$ where D_I represents the average infectious days. Parameter β is known as the effective contact rate, i.e. the average number of contacts an individual have with others is β.

In analyzing whether any pandemic is contained, it is very crucial to obtain parameter β. We next show that how we can derive β from the aforementioned differential equations.

3.1 Obtaining β_t and R_t for the SIR Model

In the SIR model, we can express the number of susceptible individuals in terms of population size and the number of infected persons as $S(t) \approx N - I(t)$. By replacing $S(t)$ with $N - I(t)$ in the second differential equation of (1), we would have

$$\frac{dI(t)}{dt} = \beta \frac{\Big(N - I(t)\Big)I(t)}{N} - \sigma I(t). \tag{2}$$

We can rewrite (2) as follows:

$$\frac{dI(t)}{(\beta - \sigma)I(t) - \frac{\beta}{N} I^2(t)} = dt. \tag{3}$$

By taking definite integral from time t_1 to t_2 and assuming β to be constant in this time interval, we would have

$$\int_{t_1}^{t_2} \frac{dI(t)}{(\beta - \sigma)I(t) - \frac{\beta}{N} I^2(t)} = \int_{t_1}^{t_2} dt \tag{4}$$

which leads to

$$\frac{1}{\beta - \sigma}\Big(\log \frac{I(t_2)}{\beta - \sigma - \frac{\beta}{N} I(t_2)} - \log \frac{I(t_1)}{\beta - \sigma - \frac{\beta}{N} I(t_1)}\Big) = t_2 - t_1 \tag{5}$$

One can easily check (5) has a unique solution for β due to the fact that term $\frac{1}{\beta - \sigma}$ and log term have monotonic behaviors.

An epidemic happens in case of increase in the number of infected individuals, i.e. $\frac{dI(t)}{dt} > 0$, or consequently

$$\beta \frac{\Big(N - I(t)\Big)I(t)}{N} - \sigma I(t) > 0. \tag{6}$$

In the early stage of an epidemic, almost everyone are susceptible except very few cases. Therefore, $N - I(t) \approx N$ and as a result, condition (6) would turn into $\frac{\beta}{\sigma} > 1$.

The variable $R \triangleq \frac{\beta}{\sigma}$ is defined as the *effective reproduction number*. It is a useful metric to determine epidemic growth. In case of having $R > 1$, the epidemic is growing exponentially while $R < 1$ indicates the epidemic is contained and will decline and die out eventually.

For discrete-time cases such as daily reporting on number of infected cases, the time-variant effective contact rate β_t, which represents the contact rate for time slot t can be derived by solving the following equation:

$$\frac{1}{\beta_t - \sigma}\Big(\log \frac{I(t+1)}{\beta_t - \sigma - \frac{\beta_t}{N} I(t+1)} - \log \frac{I(t)}{\beta_t - \sigma - \frac{\beta_t}{N} I(t)}\Big) = 1 \ \forall t. \tag{7}$$

Therefore, the time-variant effective reproduction number would be defined as $R_t \triangleq \frac{\beta_t}{\sigma}$. Since it is difficult to write a closed form solution for β_t in (7), we take a simpler approximation to β_t by considering the following which is based on (2)

$$\beta_t \approx \frac{\sigma I(t) + \Big(I(t+1) - I(t)\Big)}{(1 - \frac{I(t)}{N})I(t)}. \tag{8}$$

Then, we estimate R_t as $\frac{\beta_t}{\sigma}$.

3.2 Obtaining the Confidence Interval for R_t

Since there is uncertainty about parameter D_I (or equivalently σ) and the number of infected cases $I(t)$, we now provide the derivation of confidence interval for parameter R_t. Regarding modeling

the ambiguity in the number of the infected cases, we present the uncertainty about the actual number of infected cases as a factor of reported ones, i.e. $I_{rep}(t) \triangleq \frac{1}{K} I(t)$, and K is a constant greater than 1. The main intuition behind this factor is due to taking into account the following two phenomena, namely lack of sufficient number of tests (specially in the beginning of the pandemic) and asymptomatic cases (mild infections which might not even be noticed). To derive the confidence interval, we need to first find the marginal distribution of R_t. By considering $f_D(d)$ and $f_K(k)$ as the probability distribution function (pdf) for parameters D_I and K, respectively, the joint pdf of these parameters would be

$$f_{D,K}(d,k) = f_D(d) f_K(k) \tag{9}$$

due to the independence of D_I and K. We can derive the probability distribution function of R_t by performing the following transformation on parameters D_I and K and introducing auxiliary variable Z:

$$Z \triangleq K\,,\ R_t \quad = \frac{1}{1 - \frac{K I_{rep}(t)}{N}} \Big(1 + D_I \frac{I_{rep}(t+1) - I_{rep}(t)}{I_{rep}(t)}\Big). \tag{10}$$

Since the transformation of (Z, R_t) to (D_I, K) is one-to-one, we have

$$K = Z\,,\ D_I \quad = \frac{R_t(1 - Z a_t) - 1}{b_t}, \tag{11}$$

where $a_t \triangleq \frac{I_{rep}(t)}{N}$ and $b_t \triangleq \frac{I_{rep}(t+1) - I_{rep}(t)}{I_{rep}(t)}$, the joint pdf of Z and R_t would be $f_{Z,R_t}(z,r) = |J| f_{D,K}(d,k)$ with

$$J \triangleq \begin{bmatrix} \frac{\partial d}{\partial z} & \frac{\partial d}{\partial r} \\ \frac{\partial k}{\partial z} & \frac{\partial k}{\partial r} \end{bmatrix}. \tag{12}$$

By substituting the corresponding values of parameters and the Jacobian, we have:

$$f_{Z,R_t}(z,r) = |\frac{1 - z a_t}{b_t}| f_D(\frac{r(1 - z a_t) - 1}{b_t}) f_K(z). \tag{13}$$

The marginal pdf of R_t can be obtained by taking integral of (13) over parameter z, i.e.

$$f_{R_t}(r) = \int f_{Z,R_t}(z,r) dz = \int |\frac{1 - z a_t}{b_t}| f_D(\frac{r(1 - z a_t) - 1}{b_t}) f_K(z) dz. \tag{14}$$

Remark 1: Based on statistical experiments, one reasonable assumption regarding the pdf of parameters D_I and K is that both of them have Gaussian distributions. By considering $D_I \sim \mathcal{N}(\mu_D, \sigma_D^2)$ and $K \sim \mathcal{N}(\mu_K, \sigma_K^2)$, the pdf of R_t can be simplified as

$$\begin{aligned} f_{R_t}(r) = &\int_{-\infty}^{\frac{1}{a_t}} (\beta_0 + \beta_1 z) C \sqrt{2\pi\sigma_c^2} \phi_{\mu_c,\sigma_c^2}(z) dz \\ &+ \int_{\frac{1}{a_t}}^{\infty} (-\beta_0 - \beta_1 z) C \sqrt{2\pi\sigma_c^2} \phi_{\mu_c,\sigma_c^2}(z) dz, \end{aligned} \tag{15}$$

where $\phi_{\mu_c,\sigma_c^2}(.)$ indicates the pdf of a normal distribution with mean μ_c and variance σ_c^2 while

$$\begin{gathered} \beta_0 \triangleq \frac{1}{b_t}\,,\ \beta_1 \triangleq \frac{-a_t}{b_t}, \\ \alpha_0 \triangleq \frac{(\frac{r-1}{b_t} - \mu_D)^2}{2\sigma_D^2} + \frac{\mu_K^2}{2\sigma_K^2}\,,\ \alpha_1 \triangleq \frac{(-\frac{r a_t}{b_t})(\frac{r-1}{b_t} - \mu_D)}{\sigma_D^2} - \frac{\mu_K}{\sigma_K^2}, \\ \alpha_2 \triangleq \frac{(\frac{r a_t}{b_t})^2}{2\sigma_D^2} + \frac{1}{2\sigma_K^2}\,,\ \mu_c \triangleq \frac{-\alpha_1}{2\alpha_2}\,,\ \sigma_c^2 \triangleq \frac{1}{2\alpha_2}\,,\ C \triangleq \frac{e^{-(\alpha_0 - \frac{\alpha_1}{4\alpha_2})}}{2\pi\sigma_D\sigma_K}. \end{gathered} \tag{16}$$

By taking integral through using change of parameters, (15) can be rewritten as follows

$$\begin{aligned} f_{R_t}(r) = &-2C\beta_1\sigma_c^2 e^{-\frac{(\frac{1}{a_t} - \mu_c)^2}{2\sigma_c^2}} + C\sqrt{2\pi\sigma_c^2}(\beta_1\mu_c + \beta_0)\Phi_{\mu_c,\sigma_c^2}(\frac{1}{a_t}) \\ &+ C\sqrt{2\pi\sigma_c^2}(-\beta_1\mu_c - \beta_0)(1 - \Phi_{\mu_c,\sigma_c^2}(\frac{1}{a_t})), \end{aligned} \tag{17}$$

where $\Phi_{\mu_c,\sigma_c^2}(.)$ represents the cumulative distribution function (cdf) of a normal distribution with mean μ_c and variance σ_c^2.

The confidence interval would belong to $(\bar{R}_t - \delta, \bar{R}_t + \delta)$ where $\bar{R}_t \triangleq \mathbb{E}[R_t] = \int r f_{R_t}(r) dr$ and δ can be derived by satisfying $\mathbb{P}(|R_t - \bar{R}_t| \le \delta) = \int_{\bar{R}_t - \delta}^{\bar{R}_t + \delta} f_{R_t}(x) dx = 1 - \epsilon$ for some small $\epsilon > 0$.

3.3 Estimating the Risk Score

We propose a novel risk score metric for a given community that is proportional to the probability of someone in that community becoming infected in the next time period (typically, 24 hours). The risk score can be derived as the average number of people in that community that are likely to get infected in the next 24 hours by the currently infectious people divided by the current number of susceptible individuals. We further normalize this probability by multiplying by 10,000, so that a score of 1 implies a 1 in 10,000 chance of getting infected, a score of 2 implies a 2 in 10,000 chance of getting infected, and so on. Mathematically, the risk score is defined as follows:

$$\Gamma_t = \frac{I(t) \cdot R_t}{D_I \cdot S(t)} \cdot 10000 \approx \frac{K \cdot I_{rep,new}(t) \cdot R_t}{N} \cdot 10000, \tag{18}$$

where $I_{rep,new}(t)$ indicates the most recently reported count of new confirmed infectious cases, K refers to the ratio of true cases to reported cases, R_t is the time-varying reproduction number, and N is the total population size of the community. The approximation follows from the fact that $I_{rep,new}(t)$ is approximately equal to $\frac{I(t)}{D_I \cdot K}$ and $S(t)$ the number of susceptible people in the community is approximately equal to N in the early stages of the epidemic. Confidence intervals for the risk score Γ_t could be obtained numerically using a similar process as described for R_t accounting also for uncertainty in K. Note that since K may not be known for a given community, it may be helpful to use the following normalized form of the risk score: $\frac{\Gamma_t}{K}$, which is still proportional to the probability of infection for an individual.

3.4 Color-coded Risk Levels

To further simplify the presentation of the risk score to a wider audience, we propose to classify the risk levels into four color-coded levels: (Green, Yellow, Orange, Red). The risk level is determined by evaluating the normalized risk score ($\frac{\Gamma}{K}$) with respect to three pre-specified threshold levels $\theta_1, \theta_2, \theta_3$, such that when $\frac{\Gamma}{K} < \theta_1$ the risk level is green, when $\theta_1 \le \frac{\Gamma}{K} < \theta_2$ the risk level is yellow, when $\theta_2 \le \frac{\Gamma}{K} < \theta_3$ the risk level is orange, and when $\frac{\Gamma}{K} \ge \theta_3$ the risk level is red.

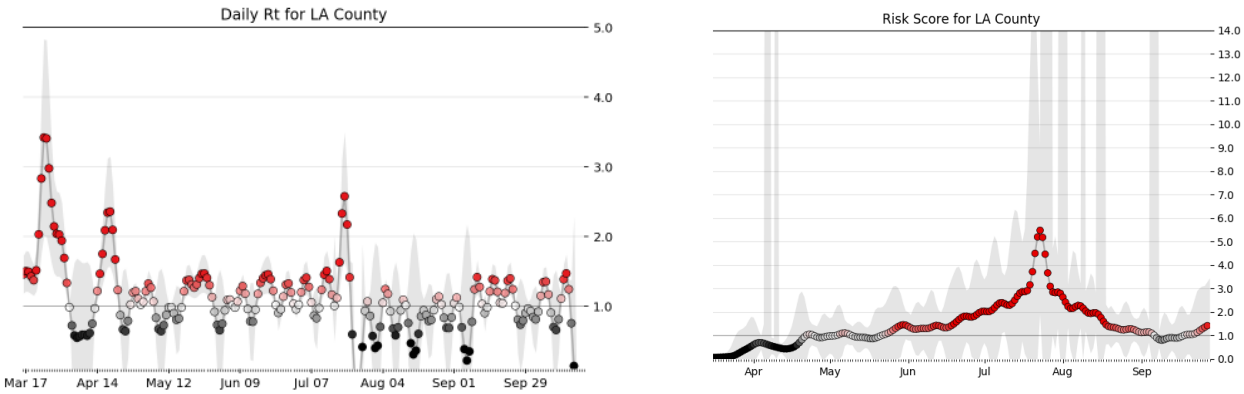

Figure 2: The left and right plots respectively represent the estimated effective reproduction number R_t and the risk score Γ_t over time for the entire county of LA considering $\mathbb{E}[D_I] = 7.5$, $Var[D_I] = 9$, $\mathbb{E}[K] = 3$, and $Var[K] = 0.44$. The gray area represents the 95% confidence interval in the estimates.

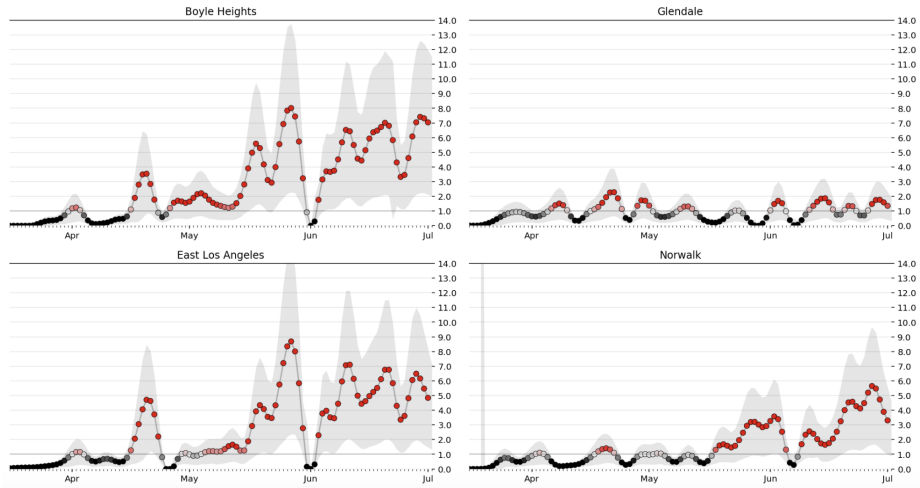

Figure 3: Estimate of risk score Γ_t over time for four representative communities in LA County: Boyle Heights, Glendale, East LA, and Norwalk. Regarding the settings, we considered the following $\mathbb{E}[D_I] = 7.5$, $Var[D_I] = 9$, $\mathbb{E}[K] = 3$, and $Var[K] = 0.44$. Our approach also yields uncertainty in the estimate, as shown in the form of confidence intervals (in gray).

4 IMPLEMENTATION AND EVALUATION IN LOS ANGELES COUNTY

The software for data collection, infection rate estimation and prediction has already been implemented and made available as open-source software (at the following repository: https://github.com/ANRGUSC/covid19_risk_estimation). The software is written in Python using standard data processing libraries such as NumPy and SciPy.

4.1 Data Sources

We have acquired COVID-19 case data from the LA County's Department of Public Health using a Python-based data parser we wrote (open-sourced at the following link: https://github.com/ANRGUSC/lacounty_covid19_data). We have been updating this repository regularly with the latest data every day since mid-march and also making available plots of the number of cases, number of fatalities, top 6 communities with the large number of cases, infection rate for the entire LA County, and the top 9 communities with the highest infection rate at the following link: http://anrg.usc.edu/www/covid19.html.

The following data sources are used for the infection rate and prediction:

- The COVID-19 case information was collected through LA County's daily press releases (Accessible through the following website: http://publichealth.lacounty.gov/media/Coronavirus/).
- Recovery information provided by the World Health Organization.

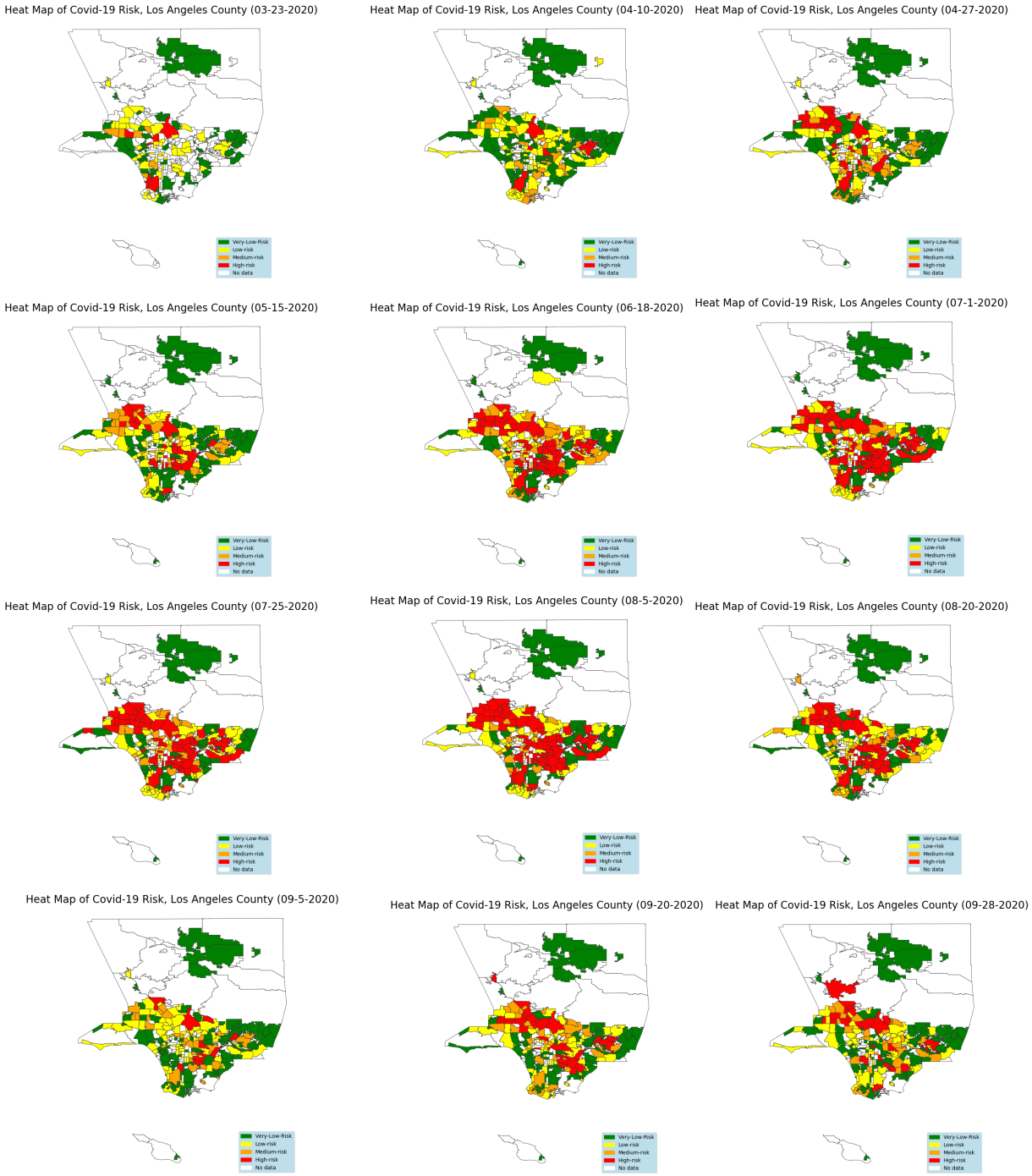

Figure 4: Maps showing the estimated risk score for different LA County Communities on different dates since mid-March 2020. Top row: March/23/2020, April/10/2020, April/27/2020; Second row: May/15/2020, June/18/2020, July/1/2020. Third row: July/25/2020, August/05/2020, August/20/2020. Last row: September/05/2020, September/20/2020, September/28/2020.

- The population data from LA County Census is available online (from lacounty.gov/government/geography-statistics/cities-and-communities/).

4.2 Real-world Usage

The City of Los Angeles is currently using the risk model described in this work that has been developed by researchers at USC, to help assess location-based risk for COVID-19 infection. The City is working with the County and other partners to develop a tool that is publicly accessible and can be used by individuals and communities to mitigate risk of infection. The goal is to change behaviors to reduce risk of infection and promote a greater understanding of factors that increase COVID risk. A color-coded COVID-19 threat level tool that can be used by citizens has also been unveiled by the Mayor of the City of LA, online at https://corona-virus.la/covid-19-threat-level. Besides, our risk score is made available to community members here https://grmds.org/risk/, wherein the users can enter their community name to understand their community's risk level. This prototype mapping site is helpful in developing options that are useful for people and continues to evolve.

5 EVALUATION RESULTS

We present plots from our analysis of LA County community case data using the estimation approach described in this work in Figure 2, Figure 3 and Figure 4. The code and the data used for generating these plots are available at the following GitHub repositories:

- **Risk model software:** - https://github.com/ANRGUSC/covid19_risk_estimation
- **COVID-19 case data for LA County:** - https://github.com/ANRGUSC/lacounty_covid19_data

Figure 2 shows plots of the estimated expected reproductive number R_t and the estimated risk score for the entire LA county. These plots are based on a 14-day moving average applied on the daily number of confirmed cases. In accordance with LA county daily press releases, there is a sharp jump in both R_t and risk score around the beginning of July. Note that the reason the risk score during the beginning of July is higher than the risk score during the last week of March, despite having the same R_t, is due to the fact that there are significantly more confirmed cases in July compared to March.

Figure 3 shows the risk score estimates over time for four representative communities within LA County. The case data continue to increase for some communities, while the number of cases remained somewhat study for a few communities.

Figure 4 shows the color-coded risk levels for communities in LA County for select dates over the past six months. The risk levels have gone up for many communities in the last week of July and the beginning of August, which is also visible in the county-wide risk score, as shown in Figure 2.

6 CONCLUSION

We have proposed a new risk metric Γ_t that can be used by individuals in any community to assess their probability of getting infected by COVID-19. The metric builds on the estimation of R_t, the average reproductive number, which is obtained from a time-varying extension of the classical SIR model. We show how to evaluate the uncertainty in both metrics as well. In future work, we plan to generalize the approach to the SEIR model, which also models an additional incubation period. We have released code to implement an estimation of the risk score that can be used for any community worldwide as long as time-series data for confirmed new cases and the population are known. We have also proposed the use of simple color-coded risk levels to inform and guide the public, as has been adopted in the City of Los Angeles. One open question to be investigated in future work is how residents respond to these color-coded levels and how to communicate the behaviors appropriate to each level.

7 ACKNOWLEDGMENTS

This work is supported by the USC Viterbi Center for Cyber-Physical Systems and the Internet of Things (CCI).

REFERENCES

[1] Marshall Maximilian Dong Ensheng Squire Marietta Badr Hamada, Du Hongru and Gardner Lauren. 2020. Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *The Lancet Infectious Diseases* (2020).
[2] Luís Bettencourt and Ruy Ribeiro. 2008. Real Time Bayesian Estimation of the Epidemic Potential of Emerging Infectious Diseases. *PloS one* 3 (02 2008), e2185. https://doi.org/10.1371/journal.pone.0002185
[3] Fred Brauer. 2005. The Kermack–McKendrick epidemic model revisited. *Mathematical biosciences* 198, 2 (2005), 119–131.
[4] Sheryl L. Chang, Nathan Harding, Cameron Zachreson, Oliver M. Cliff, and Mikhail Prokopenko. 2020. Modelling transmission and control of the COVID-19 pandemic in Australia. arXiv:q-bio.PE/2003.10218
[5] Yu-Chun Chen, Ping-En Lu, and Cheng-Shang Chang. 2020. A Time-dependent SIR model for COVID-19. *ArXiv* abs/2003.00122 (2020).
[6] Kai Kupferschmidt. 2020. The lockdowns worked—but what comes next?
[7] Ying Liu, Albert A Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. 2020. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of travel medicine* (2020).
[8] M. Newman. 2010. Networks: An Introduction. In *Oxford University Press*.
[9] Kevin Systrom. 2020. The Metric We Need to Manage COVID-19. http://systrom.com/blog/the-metric-we-need-to-manage-covid-19/
[10] Aurelio Tobías. 2020. Evaluation of the lockdowns for the SARS-CoV-2 epidemic in Italy and Spain after one month follow up. *Science of The Total Environment* (2020), 138539.