

Quantifying Mobility and Mixing Propensity in the Spatiotemporal Context of a Pandemic Spread

Satyaki Roy , Preetom Biswas , and Preetam Ghosh 

Abstract—COVID-19 is the most acute global public health crisis of this century. Current trends in the global infected and death numbers suggest that human mobility leading to high social mixing are key players in infection spread, making it imperative to incorporate the spatiotemporal and mobility contexts to future prediction models. In this work, we present a generalized spatiotemporal model that quantifies the role of human social mixing propensity and mobility in pandemic spread through a composite latent factor. The proposed model calculates the exposed population count by utilizing a nonlinear least-squares optimization that exploits the intrinsic linearity in SEIR (Susceptible, Exposed, Infectious, or Recovered). We also present inverse coefficient of variation of the daily exposed curve as a measure for infection duration and spread. We carry out experiments on the mobility and COVID-19 infected and death curves of New York City to show that boroughs with high inter-zone mobility indeed exhibit synchronicity in peaks of the daily exposed curve as well as similar social mixing patterns. Furthermore, we demonstrate that several nations with high inverse coefficient of variations in daily exposed numbers are amongst the worst COVID-19 affected places. Our insights on the effects of lockdown on human mobility motivate future research in the identification of hotspots, design of intelligent mobility strategies and quarantine procedures to curb infection spread.

Index Terms—Human mobility policies, lockdown, optimization, social mixing, spatiotemporal model.

I. INTRODUCTION

THE scourge of epidemics and pandemics has been a part of human history since time immemorial. Considering the past millennium alone, right from as early as 1317, innumerable outbreaks such as plague, flu and Ebola have globally claimed millions of lives [1]. The latest addition to the list of outbreaks is COVID-19, which, since its inception in China in December 2019, has brought the world to a veritable standstill. COVID-19 has followed a similar course like the plague, flu and Ebola and claimed over 2 million lives globally as of January 2021, while its severity continues to burgeon in the US, U.K., Brazil and

parts of Asia [2] with a sizable number still projected to die in the subsequent waves of this pandemic.

Most countries reacted poorly to the looming dangers of COVID-19. In the absence of credible vaccination treatment [3], social distancing and ensuing lockdown are threatening to bring the global economy to a halt. There has been a drop in industrial productivity, stock exchange percentage and increase in the price of goods [4] as well as a potential contraction in US GDP [5]. The world is on the brink of a COVID-19-induced recession (with as many as 709 000 seeking unemployment aid in the US alone [6]). Nations are now willing to relax the lockdown to mitigate economic losses [7]. Research on clinical, epidemiological or socioeconomic implications of COVID-19 is stymied by the absence of prior knowledge [3], [8]. The reliability of the data is challenged by high variability in testing and surveillance or contact tracing-based detection of the prospective infected population. Finally, logistical factors such as dearth of or the accuracy in testing, reluctance in reporting death and recovery [9] and dubious information in print and social media [10] further misguide precautionary and mitigation measures. It is known that human mobility across neighboring geographic regions (such as imports of international travellers [11]) leading to high social contact is the primary mode for spread, yet there exists no model that quantifies the joint effect of human mobility and social mixing in the spatiotemporal dynamics of pandemic spread.

Epidemiological models such as SEIR, SIR, SEIRD, SEIRS (susceptible, exposed, infected (or infectious), recovered, or death), etc. and their variants have been employed to study the spread of infection [12], [13]. As per the susceptible exposed infected recovered (SEIR) model, the *susceptible* (S) class comprises individuals who are not exposed to the infection. Once exposed to an infected individual, the susceptible may transfer to the *exposed* (E) category. The E class represents asymptomatic or untested individuals, who transition to the (tested) *infected* (I). Individuals in I transition to the recovered (R) categories [14]. There is also the non-epidemiological modeling analysis proposed by the Institute for Health Metrics and Evaluation (IHME) [15] squarely on the basis of mortality rates. Both of these techniques have their shortcomings. With regard to stochastic epidemiological models, combinations of state transition parameters can show a good fit with the training data but yield disparate model predictions – a problem defined as parameter identifiability [16]. On the other hand, increasing efforts to trace the passage of infection from mobility statistics [17] militates against the efficacy of the IHME model that

Manuscript received November 15, 2020; revised January 30, 2021; accepted February 8, 2021. Date of publication March 17, 2021; date of current version May 25, 2021. This work was supported by NSF under Grant CBET-1802588. (Corresponding author: Satyaki Roy.)

Satyaki Roy is with the Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27514-3916 USA (e-mail: satyakir@unc.edu).

Preetom Biswas is with the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh (e-mail: preetomicc@gmail.com).

Preetam Ghosh is with the Department of Computer Science, Virginia Commonwealth University, Richmond, Virginia 23284 USA (e-mail: pghosh@vcu.edu).

Digital Object Identifier 10.1109/TETCI.2021.3059007

assumes that the mortality rate follows a normal distribution, while not factoring in effects of transmission dynamics nor contact patterns on spread.

Given the lack of prior knowledge on dealing with a public health crisis of this scale, the policymakers are ill-equipped to design mitigation strategies. To bridge the gap, the research community of epidemiologists, clinicians and computer scientists are applying their expertise to seek out factors and their effects on infection spread as well as the impending economic crash [3]. First, machine learning (ML) is helping build prediction models on epidemiological and clinical data. Given existing clinical data, prediction models [18] and therapeutic approaches can help identify vulnerable groups [19], [20]. Epidemiologists are trying to identify spread dynamics of COVID-19. Inga Holmdahl *et al.* [21] analyze the pros and cons of forecasting models that make predictions through curve fitting or mechanistic models, while supervised and unsupervised ML is helping trace the trends in infection dynamics [22]. Khan *et al.* performed regression analysis, cluster analysis and principal component analysis on Worldometer infection count data to gauge the variability and effect of testing in prediction of confirmed cases [4]. Roy *et al.* used regression analysis to pinpoint pre-lockdown factors that affect post-lockdown pandemic numbers [23].

Modeling COVID-19 using SEIR: There have been efforts to employ SEIR to study the effects of demography, immunity and social distancing on infection spread. He *et al.* employed the particle swarm optimization on the COVID-19 data of Hubei province of China to calculate the parameters of the SEIR model. They discuss how these parameters can vary with demography [24]. Pandey *et al.* utilized the SEIR model with regression on the COVID-19 data of India collected by Johns Hopkins University in the interval of 30th January to 30th March, 2020 to show the reproduction number to be approximately 2 [25]. Yang *et al.* applied artificial intelligence on the COVID-19 data of Hubei, China into the SEIR model to estimate the date when the infection peaks. They also predicted how the quarantine will affect the dynamics of contagion [12]. Annas *et al.* calculated the parameters of SEIR model by incorporating the factors of vaccination and isolation. They applied the model on the COVID-19 data of Indonesia to study the long-term effects of vaccine and isolation on curbing spread [26]. Radulaescu *et al.* adapted SEIR to study spread dynamics in an age-heterogeneous scenario. As a case study, they simulate a small community in New York and assess the effects of control measures such as restricted mobility, social distancing and lockdown [27]. Iwata performed simulation using the SEIR model to predict the effect of secondary outbreak in a community outside China. They demonstrate that the timing of hospital visits may affect the outbreak [28]. Mwalili *et al.* applied the modified SEIR model to study the effect of pathogens and intervention measures on disease spread. They discuss the ill-effects of flouting social distancing and basic hygiene measures on COVID-19 [29]. Tang *et al.* adapted the SEIR model to incorporate the assumption that the infected person may act as a vector of infection during the incubation period. They use the model to make recommendations and prediction of the disease spread [30]. Lopez *et al.* utilize the SEIR epidemic model to study the consequence of quarantine on

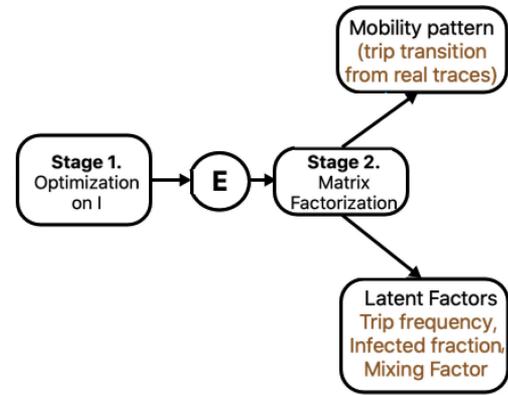


Fig. 1. Contributions of this work. First, we present an optimization that employs the daily infected (I) to infer the daily exposed (E) numbers of a region. Second, we utilize E, in combination with the mobility pattern (obtained from real human mobility traces), to calculate the latent factors for infection spread that quantifies mobility and social mixing.

the population of Spain and Italy. They show that isolation can help achieve a 10 times decline in disease spread. This has been corroborated by studying contagion in the pre- and post-COVID intervention in Italy [31].

Contributions: In this paper, we make three major contributions. First, we introduce a generalized spatiotemporal framework, the first of its kind, that quantifies the components affecting infection spread through a *latent factor*. Specifically, this latent factor is a metric quantifying the joint influence of human mobility and social mixing on the exposure to an infection (see Fig. 1). We demonstrate its efficacy by employing this spatiotemporal model on New York City mobility traces and COVID-19 data trends. Second, we argue that the extent and spread of infection can be gauged in terms of the projected exposed (i.e., asymptomatic individuals) numbers, instead of the infected and mortality count (that has been deemed a reliable measure for the extent of infection spread for a geographical region [15]). Third, we adapt a well-studied measure for dispersion in statistical distribution, called *coefficient of variation*, as a measure to quantify the potential for infection spread and duration, and demonstrate that nations with a high inverse of the coefficient of variation in daily exposed numbers are amongst the most COVID-19 affected. Finally, we discuss how the proposed spatiotemporal model can identify pandemic hotspots as well as the ideal time and extent of lockdowns to minimize contact during a pandemic. The exposed population of a region is an input to the spatiotemporal model that estimates latent factors. The proposed approach employs a nonlinear least-squares optimization to infer the daily exposed numbers. It incorporates an exposed to infection transition step of the complete SEIR (i.e., $S \rightarrow E \rightarrow I \rightarrow R$). It is important to mention here that the stated optimization is just one approach to gauge the exposed numbers and that the spatiotemporal model will work seamlessly for the exposed estimates using other approaches as well.

This paper is organized as follows. Section II introduces the major contributions of this work, namely, the optimization to estimate exposed, spatiotemporal model and inverse coefficient of variation to quantify spread. Section III presents the

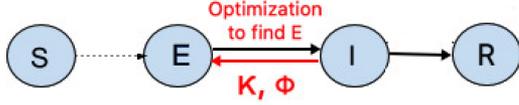


Fig. 2. State transitions in the SEIR model are shown in black arrows. The optimization to calculate the daily exposed from infected, by calculating the fraction and duration of transition from E to I (ϕ , κ , respectively) is highlighted in red.

experimental results on traffic and COVID-19 data from New York City and the world. Finally, Section IV draws the conclusions.

II. APPROACH

Susceptible Exposed Infected Recovered Death model: In the SEIR model [14], the *susceptible* (S) class comprises individuals who are not exposed to infection. Once exposed to infected individuals, they may transfer to the *exposed* (E) category. E class are asymptomatic or untested individuals, who transition to the (tested) *infected* (I). The individuals in I transition to either *recovered* (R) (or *dead*) (Fig. 2).

Estimation of daily exposed: We discuss the optimization that utilizes the daily infected numbers to estimate the daily exposed numbers (see Fig. 2). This is based on the SEIR model that states that a fraction of susceptible individuals transition to exposed on contact with infected state, while a fraction (say, ϕ) over time (say, κ days) transfer to infected. We estimate E by assuming that a mean fraction (ϕ) of $e_t \in E$ transition to the infected category in mean duration κ time. We minimize average squared error between the fraction of the predicted daily exposed at time t (i.e., $\phi \times e_t$) and infected population i_t at time $t + \kappa$ (i.e., $i_{t+\kappa}$).

$$\min_{E, s, \kappa, \phi} \frac{1}{T - \kappa} \sum_{t=1}^{T-\kappa} (\phi \times e_t - i_{t+\kappa})^2 \quad (1)$$

$$\text{s.t.} \quad 0 < \phi \leq 1.0, LB \leq \kappa \leq UB \quad (2)$$

$$0 < s \leq 1.0, e_t \leq s \times P(x) \quad (3)$$

Ex. 1 ensures that the daily exposed curve scaled by a factor ϕ and shifted by κ days on the time axis is nearly identical (i.e., having low mean squared error) to the daily infected curve. Constraint 2 causes the incubation period κ and infection rate ϕ to be in range $[LB, UB]$ and $(0, 1]$, respectively. Finally, constraint 3 ensures that, given a place x with population $P(x)$, the optimizer considers the upper bound for daily exposed e_t to be a fraction, say s , of $P(x)$. We illustrate an example in Fig. 3, where $\kappa = 14$ days and $\phi = 0.6$. Given a daily infected curve (shown in blue) that peaks on day 50, the optimizer should infer a daily exposed curve (shown in green) that shows a higher curve peaking at day $50 - \kappa = 36$.

A. Inverse Coefficient of Variation

Coefficient of variation (CV) is a statistical measure of the variability of a distribution with respect to its mean. It was conceived to compare data varying in units, say the height of

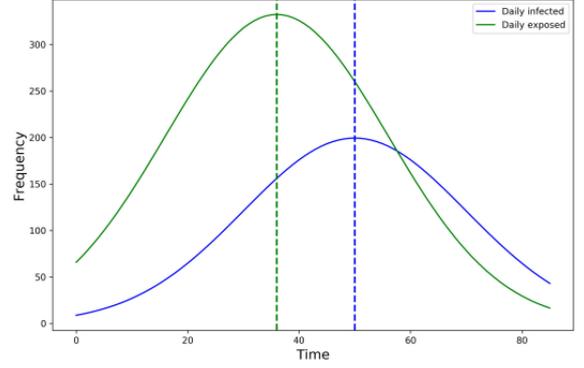


Fig. 3. Daily exposed curve (colored green) and daily infected curve (colored blue) for $\kappa = 14$ days and $\phi = 0.6$.

a child and an adult [32]. We posit that the inverse of CV (i.e., ICV) can be an effective measure for the potential threat posed by a pandemic in a geographical region. It is measured as $\frac{\mu}{\sigma}$, where μ and σ are the mean and standard deviation of an exposed curve, respectively. ICV was termed the *reward-to-variability* ratio by American economist and Nobel laureate William Sharpe and used to gauge the performance of mutual funds as a ratio between return on investment and market variability [33]. In the context of pandemic, *ICV of the daily exposed curve* quantifies the ratio between the potential for infection spread over time to its variability, suggesting that it can be an effective measure for the potential extent and duration of pandemic spread in any geographical region.

B. Spatiotemporal Modeling

We present a spatiotemporal model that helps to quantify the daily exposed numbers in terms of a latent factor combining social mixing and human mobility (refer Fig. 1). We discuss the preliminaries on matrix normalization as well as frequency and transition matrix before formalizing the model.

1) *Column Normalization of a Matrix:* Given any two-dimensional matrix $Z \in \mathbf{M}_{m \times n}(\mathbb{R})$, we define a left stochastic matrix (i.e., matrix with column summing to 1), as follows:

$$\hat{Z} = \left\{ \frac{v_{ij}}{\sum_{k=1}^m v_{kj}} : v_{ij} \in Z \right\} \quad (4)$$

2) *Frequency and Transition Matrix:* Given a geographical region with a set of geographical sub-regions (or zones) B , the frequency matrix $F \in \mathbf{M}_{|B| \times |B|}(\mathbb{R})$ is created from the human mobility traces, where $f_{i,j} \in F$ denotes the number of trips made from zone $b_j \in B$ to $b_i \in B$. We generate a transition matrix $A \in \mathbf{M}_{|B| \times |B|}$ performing column normalization of F (as defined in Section II-B1). Each element of this matrix $a_{i,j} \in A$ is the probability of making a trip from b_j to b_i . The frequency or transition matrix captures the overall mobility trends within and across zones of any given geographical region.

a) *Quantifying trip count:* In keeping with Markov chain, we calculate the c -th power of A that represents the probability of transitioning from one zone to another in exactly c trips [34]. Given that $A_{i,j}^c = \sum_{k=1}^{|B|} a_{i,k}^{c-1} a_{k,j}$, the (i, j) -th entry in A raised

to power 2 can be written as:

$$a_{i,j}^2 = a_{i,1}.a_{1,j} + a_{i,2}.a_{2,j} + a_{i,3}.a_{3,j} + \dots \quad (5)$$

If $i = 1$, $j = 2$ and $z = 2$ in (5), we obtain the likelihood for a trip from b_2 to b_1 in 2 hops (with b_k ($1 \leq k \leq |B|$) as an intermediate stop). In (5), the term $a_{1,1}.a_{1,2}$ is the probability of traveling from b_2 to b_1 in the order $b_2 \rightarrow b_1 \rightarrow b_1$; analogously, $a_{1,2}.a_{2,2}$ is the probability of commute in the following order $b_2 \rightarrow b_2 \rightarrow b_1$, and so on.

Let us assume that an individual makes $1, 2 \dots, r$ trips. We calculate the stochastic matrix corresponding to the inter- and intra-zone transition for less than or equal to r trips ($\hat{A}_{\leq r}$), where $A_{\leq r}$ is defined as:

$$A_{\leq r} = \sum_{i=1}^r A^i \quad (6)$$

We assume that each trip length is independent of another. In other words, an individual can independently choose to make i , $i + 1$ or more trips across different zones in a region.

3) *Formal Definition:* Let $\gamma_{bt} \in \Gamma$ be the number of daily exposed individuals at zone $b \in B$ at time t and \hat{A} (i.e., $\hat{A}_{\leq r}$ for some $r \in \mathbb{Z}$) defined in Section II-B2) be the transition matrix. We define the relationship $\hat{A}.X = \Gamma$, written as:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,|B|} \\ a_{21} & \dots & & \\ \vdots & & & \\ a_{|B|,1} & & & \end{bmatrix} \times \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,t} \\ x_{21} & \dots & & \\ \vdots & & & \\ x_{|B|,1} & & & \end{bmatrix} = \begin{bmatrix} \gamma_{1,1} & \gamma_{1,2} & \dots & \gamma_{1,t} \\ \gamma_{21} & \dots & & \\ \vdots & & & \\ \gamma_{|B|,1} & & & \end{bmatrix}$$

Explanation: Recall from the discussion on the SEIR model in Section I, the susceptible (S) population contract the infection via contact with the infected (I) individuals. X is the composite combination of mobility and mixing among the S and I population over time, and \hat{A} controls the extent of contact among S and I due to intra- and inter-zone mobility, resulting in the generation of the final matrix of exposed individuals over time Γ . The other features are summarized below.

- Latent factor X is a unified metric for mobility and social mixing and an element $\gamma_{i,k} \in \Gamma$ can be calculated as:

$$\gamma_{i,k} = a_{i,1}.x_{1,k} + a_{i,2}.x_{2,k} + \dots = \sum_{j=1}^{|B|} a_{i,j}.x_{j,k} \quad (7)$$

- Recall that the frequency of trips from b_i to b_j may be inferred from element $f_{j,i}$ in frequency matrix F (defined in Section II-B2). For each b_i , we calculate the trip frequency factor (ρ_i) as the total number of trips made from b_i to all boroughs including itself (i.e., $\rho_i = \sum_{j=1}^{|B|} f_{j,i}$).
- We posit that an element of the latent factor $x_{i,t} \in X$ is a combination of the three factors of a borough b_i : (a)

frequency of trips made by b_i (α_i), (b) fraction of infected individuals in b_i ($l_{i,t}$), and (c) intra- and inter-borough mixing of b_i . Thus, $x_{i,t}$ can be written as:

$$x_{i,t} = \alpha_i \times l_{i,t} \times \pi_{i,t} = \left(\frac{\rho_i}{\sum_{j=1}^{|B|} \rho_j} \times w \right) \times \frac{I_{i,t}}{P_i - D_{i,t}^C - R_{i,t}^C} \times \pi_{i,t} \quad (8)$$

Here, w is the long-term mean trip count made within and across boroughs and P_i is the population of b_i . The first term α_i is a measure of expected number of trips starting at b_i at time t ; the second term l_i is the ratio between the number of infected people at time t , $I_{i,t}$, and the number of people in borough i (barring cumulative recovered $R_{i,t}^C$ and dead $D_{i,t}^C$). The third term $\pi_{i,t}$ is the mixing factors that account for several region-specific parameters, such as susceptible count, testing frequency, strain of infection, immunity acquired against infection, etc.

4) *Modeling Lockdown:* Lockdown is modeled as restricted mobility achieved by scaling down the frequency of trips made by borough i (α_i). Given a lockdown rate be η , we achieve trip minimization, by simply scaling each element of latent factor matrix X by η , as shown in the equation below.

$$X'_{i,t} = X_{i,t} * \eta \quad (9)$$

$$\Rightarrow X'_{i,t} = (\alpha_i \times l_{i,t} \times \pi_{i,t}) \times \eta \quad (10)$$

$$\Rightarrow X'_{i,t} = (\alpha_i \times \eta) \times l_{i,t} \times \pi_{i,t} \quad (11)$$

$$\Rightarrow X'_{i,t} = \alpha'_i \times l_{i,t} \times \pi_{i,t} \quad (12)$$

In the above equation, α'_i is the scaled down frequency of trips. Note that the drop in exposed numbers is commensurate with the decrease in η , since $\gamma'_{i,k}$ in (7) can be written as $a_{i,1}.\eta.x_{1,k} + a_{i,2}.\eta.x_{2,k} + \dots = \sum_{j=1}^{|B|} a_{i,j}.\eta.x_{j,k} = \eta \times \sum_{j=1}^{|B|} a_{i,j}.x_{j,k} = \gamma_{i,k} \times \eta$. However, the knowledge of the transition matrix \hat{A} and latent factor X can allow us to devise more intelligent lockdown strategies. In the experimental results (Section III-C3) we consider a scenario where, instead of a uniform lockdown rate η , lockdown levels can vary over time (i.e., η_t). Since the magnitude of elements in \hat{A} and X vary across boroughs and time and lockdown entails economic losses, it is possible to utilize the latent factor matrix to balance joint goals of minimizing exposure and economic losses.

5) *Determination of Latent Factors:* Given transition and exposed matrices \hat{A} and Γ , we solve for the latent factor matrix X , while constraining $x_{i,t} \in X$ to be positive real numbers.

$$\min_X \hat{A}.X - \Gamma \quad (13)$$

$$\text{s.t. } x_{i,t} > 0 \quad \forall x_{i,t} \in X \quad (14)$$

C. Mean-Centered Cosine Similarity

We estimate the similarity between two vectors \mathbf{v}_1 and \mathbf{v}_2 using the cosine index of mean-centered vectors $\mathbf{v}'_i = \mathbf{v}_i - \mu_{\mathbf{v}_i}$ that measures the cosine of angle between vectors v_1 and

TABLE I
LIST OF PARAMETERS AND THEIR VALUES

Parameter	Notation	Value
Upper bound for daily exposed (Eq. 1)	s	0.3
Lower bound of incubation period (κ)		2
Upper bound of incubation period (κ)		30
Savitzky–Golay window size	-	31
Savitzky–Golay window order	-	3
Average number of trips per day	w	3×10^5
Transition matrix threshold	th	10^{-10}

v_2 as $\cos(\theta) = \frac{\mathbf{v}'_1 \cdot \mathbf{v}'_2}{\|\mathbf{v}'_1\| \cdot \|\mathbf{v}'_2\|}$. Mean-centering is a standard practice in statistical models and data-driven recommendation systems [35], [36] that allows comparison of data with varying orders of magnitude.

III. EXPERIMENTAL RESULTS

The results are classified into four subsections: (A) parameter identification and quantification of pandemic spread, (B) mobility patterns within and across zones in a region, (C) influence of mobility and spatiotemporal mixing on pandemic spread and (D) exploratory analysis. Simulation parameters are summarized in Table I. We consider the incubation period $\kappa \in [2, 30]$. Although symptoms show up in about 5 days after contact, symptoms have also been reported to appear as early 2 days after exposure [37]. For 10% of the population, the incubation period was longer than 2 weeks and, in a few cases, more than 20 days [38]. This period can potentially be extended due to delays and inaccuracies in testing.

Data collection: We discuss the NYC map and mobility traces and the (NYC and global) infected and death numbers.

1) *Map Generation and Location Identification:* The list of NYC boroughs and districts is extracted from Wikipedia [39], and the latitude and longitude of the 5 boroughs and 59 districts are taken from the Python library for geocoding services, called GeoPy [40]. The distance between any pair of points (i.e., boroughs or districts) on the NYC map is calculated using the geodesic distance function of GeoPy.

2) *NYC Mobility Data:* We source the mobility data of NYC traffic from NYCOpenData [41] – a data repository for fields ranging from city government, education, environment, health to public safety, recreation, social services and transportation. The stated data (spanning a period from 2014 to 2019), collected by the Department of Transportation of New York Metropolitan Transportation Council (NYMTC), has following fields: *ID*, *road name*, *source and destination intersecting street name*, *compass direction*, *date* and *time*. We use this data to calculate the transition matrix (see Section II-B) that captures the probability of travelling within and across boroughs.

3) *Cumulative Daily Infected and Death for NYC:* We collect COVID-19 daily infected numbers from the website of the NYC Department of Health and Mental Hygiene repository [42] that contains the data on Coronavirus Disease 2019 (COVID-19)

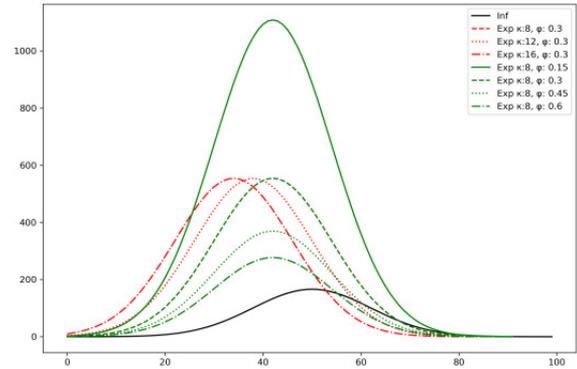


Fig. 4. Predicted exposed for varying κ and ϕ values.

in New York City (NYC). The data spans a period starting March, 2020 (which happens to be the date of first documented laboratory-confirmed cases) to November 2020.

4) *Global Cumulative Daily Infected and Death:* The time-series data of the world daily infected and death numbers is sourced from the World Health Organization, over a period spanning January 03, 2020 - October 23, 2020 [43].

A. Parameter Identification and Spread Quantification

We estimate the zone-specific parameters of infection spread (i.e., κ and ϕ) for countries and quantify the duration and spread of infection using the inverse coefficient of variation.

1) *Effect of κ and ϕ :* For a fixed infected curve (black curve), we study the variation in exposed curve with varying rate parameters ϕ controlling the fraction of population transitioning from exposed to infected and delay parameter (in days) κ (Fig. 4). For $\phi = 0.15$ and $\phi = 0.6$, the smallest and largest fraction of exposed individuals (shown in green) transition to infected, while $\kappa = 4$ and $\kappa = 16$ (red curve) cause the lowest and highest delay in exposed to infected transition respectively.

2) *Spatial Context in Global Infection Spread:* We utilize the global COVID-19 infected and death numbers (discussed in Section III-4) to estimate κ , ϕ values as well as the daily exposed and recovered numbers (as per the optimization discussed in Section II). Parameters for the select 20 countries are enlisted in Table II. Fig. 5 depicts each country in a different color and the day in the observed 300-day period when its projected daily exposed numbers peak. There are considerable variations in daily infected (and consequently exposed) numbers, as illustrated by the exposed curves of China and USA in Fig. 6(a). It is worth noting that several countries in close proximity, such as (Group 1) Iran, Iraq UAE and India (shown in red dotted circle) and (Group 2) Italy, Belgium, Germany, Austria and Romania (shown in blue dotted circle) peak nearly at the same time (see Fig. 5), alluding to the fact that mobility across neighboring zones oftentimes plays a role in a pandemic spread and affect the timing of exposed (and infected) peaks.

3) *Quantification of Infection Spread:* In addition to high population density, variations in κ and ϕ affect the extent and rate of transition from exposed to infected states. We attempt to quantify this dynamic of spread using the inverse coefficient of

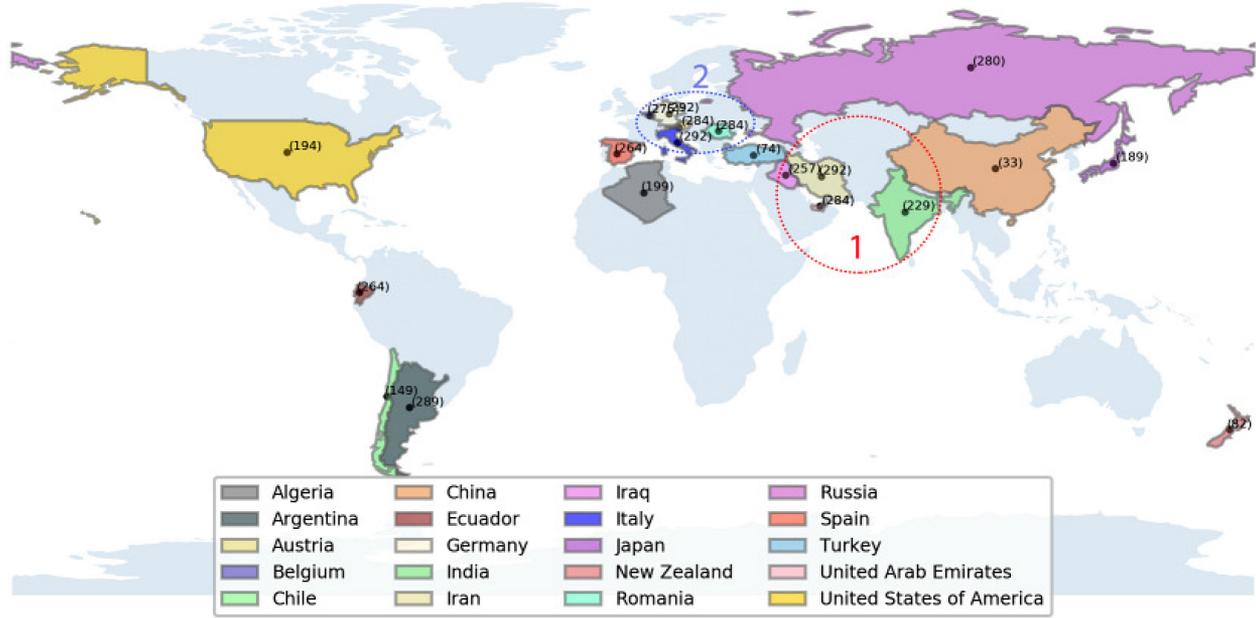


Fig. 5. Days for the exposed curve to peak for 20 countries, where each country is shown in a different color and annotated by the day in the observed 300-day period when its projected daily exposed numbers peak. There are two groups of countries (marked in red and blue dotted circles, respectively) in close proximity where the exposed numbers peak at the same time.

TABLE II
OPTIMIZATION PARAMETERS κ AND ϕ CORRESPONDING TO THE DIFFERENT COUNTRIES, ALONG WITH GOODNESS OF FIT R^2 SCORE

Country	(κ, ϕ)	R^2
Algeria	5, 1.0	0.99
Argentina	5, 1.0	1.0
Austria	10, 0.99	0.99
Belgium	14, 0.98	1.0
Chile	14, 0.69	0.99
China	5, 1.0	0.99
Ecuador	2, 0.99	0.76
Germany	2, 0.99	1.0
India	30, 0.87	0.99
Iran	2, 1.0	0.99
Iraq	2, 0.99	1.0
Italy	2, 0.99	0.99
Japan	30, 0.98	1.0
New Zealand	5, 1.0	0.99
Romania	10, 1.0	0.99
Russia	14, 1.0	0.99
Spain	30, 0.99	0.99
Turkey	30, 0.93	0.99
UAE	10, 1.0	1.0
USA	10, 0.74	1.0

variation (ICV) (defined in Section II-A) of the exposed curve. This is because, high ICV of the daily exposed curve for any given region implies a high μ (i.e., high exposed numbers) or low σ (i.e., steady exposed numbers), or both. For instance, in Fig. 6(a), the high mean exposed counts of USA contribute to its high ICV; while Iran, despite having 25% of the population of USA, has a steady (i.e., low standard deviation in the) daily exposed curve. In Table II, we summarize κ and ϕ of the

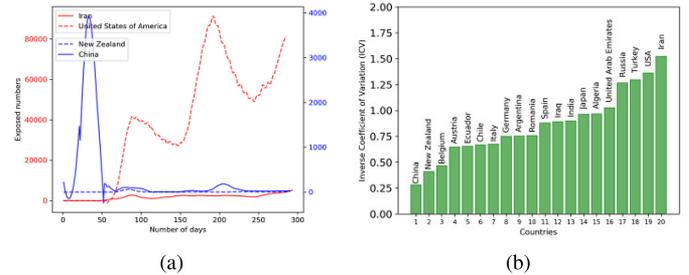


Fig. 6. Quantifying infection. (a) daily exposed of the two countries with the highest ICV (USA and Iran) and lowest ICV (China and New Zealand) smoothed using Savitzky-Golay filter, (b) inverse coefficient of variation for 20 countries.

20 countries, along with the goodness of fit R^2 for the least squared optimization (see Expression 1). It is noteworthy that inverse coefficient of variation (ICV) is useful particularly when the available time-series data covers a considerable duration, allowing for the curve to reach its first major peak within the data collection period. If the exposed curve peaks towards the end, we see near-exponential growth, resulting in high σ and low ICV. In Fig. 6(b), we plot the ICV for the 20 nations, where China, New Zealand and USA, Iran have the least ICV and highest ICV, respectively. Reports corroborate these numbers, suggesting that ICV is indeed a reliable measure of infection duration. Though the earliest cases of COVID-19 was reported in China, the nation prides itself on curbing spread by enforcing the strictest lockdown measures [44]. New Zealand has a similar story of becoming the “emblematic champion of proper prevention” due to smart and early intervention measures [45]. On the other hand, USA continues to register record new cases

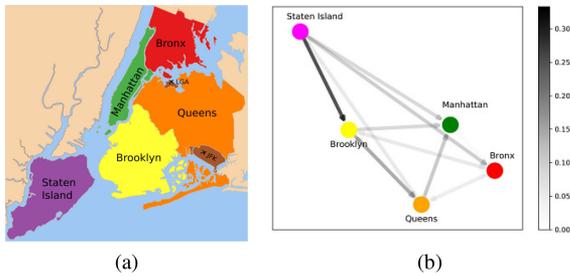


Fig. 7. Mobility pattern: (a) borough map of NYC, and (b) directed graph representation of the boroughs and mobility pattern of NYC; large circles are boroughs marked by the respective colors. The size of a borough node is proportional to the frequency of intra-borough trips, and the opaqueness of the directed edge (u, v) is proportional to the propensity of trips made from borough u to borough v .

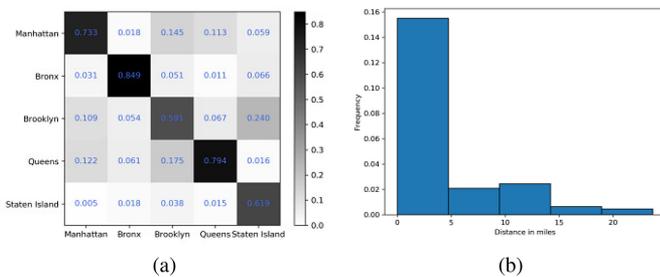


Fig. 8. Spatial context in mobility of NYC. (a) Heatmap showing the transition matrix A , where $a_{i,j} \in A$ is the probability of moving from borough j to i (written in blue), (b) histogram of bin-size 5 showing the relationship between frequency of trips made and corresponding distances in miles.

which are projected to grow in the months to come [46]. Iran too has reported unprecedented growth in new cases in October 2020 [47].

B. Spatial Context to Human Mobility Patterns

We carry out a case study on the mobility pattern of NYC and its implications on any pandemic spread. Fig. 7(a) shows the 5 boroughs of NYC. We process the human mobility data of NYC (discussed in Section III-2) to generate the frequency matrix (F) and represent the mobility within and across boroughs in a directed graph in Fig. 7(b). Each borough and district is placed according to its latitude-longitude coordinates and the size of the borough nodes and the opaqueness of a directed edge (u, v) are proportional to the fraction of total trips originating at borough u that have a destination borough v .

Fig. 7(b) shows that Staten Island to Brooklyn, followed by Brooklyn to Queens exhibit the highest inter-borough mobility. Fig. 8(a) is the transition matrix (A) from column borough to row borough labeled by the corresponding transition probabilities (discussed in Section II-B2) in the form of a heatmap, showing that intra-borough trips outnumber inter-borough trips for all boroughs. Fig. 8(b) is frequency plot of NYC trips against the distance (in miles) between the source and destination zones, where short trips are preferred over long trips.

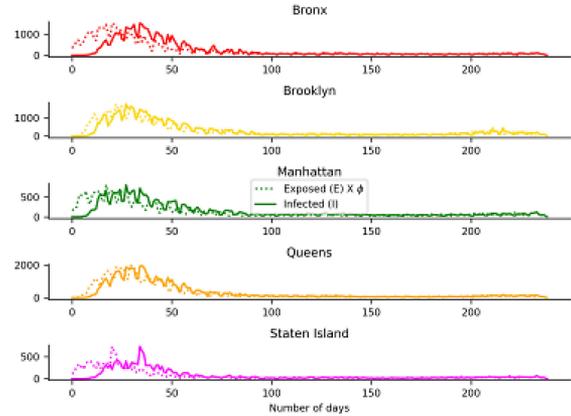


Fig. 9. Comparison of predicted daily exposed $E \times \phi$ and I for Manhattan, Bronx, Brooklyn, Queens and Staten Island.

a) Factors affecting human mobility: Human mobility is a combination of several deterministic and non-deterministic factors such as intent, convenience, environmental constraints, and so on. There are pedestrian based mobility models, such as Least Action Trip Planning [48], that suggest that a person chooses a destination (called *waypoint*) close to its current position, while another mobility framework called ORBIT [49] suggests that individuals cyclically move from one predetermined hub to another (as illustrated in Figs. 7(b), 8(a) and 8(b)). Social network-based mobility models, such as Social Network Theoretical (SNT) [50], suggest that people preferentially select next stops based on social affinity, such as work, social ties or friendships. Note that there are factors besides distance, such as intent (this can be a function of occupation, social affinity, etc.) that determine inter and intra-zone trips. Thus, despite high distance, there are a high number of trips made from Staten Island to Brooklyn and from Brooklyn to Queens. However, mobility (based on intent or proximity) across neighboring zones affect social mixing.

C. Spatiotemporal Model for Pandemic Spread

Based on the infected data (see Section III-3), we solve the optimization problem (Expression 1) to estimate the daily exposed population count. We use the Python SciPy *differential evolution* solver [51] that stochastically finds the minima by searching large areas of the candidate space. Fig. 9 shows the comparison of predicted daily exposed (dotted) E scaled down by the infection rate ϕ against that of daily infected (solid) I curve, while the lags in the corresponding peaks of the E and I curves capture the incubation period κ for a borough.

Observe that Brooklyn and Queens, the boroughs with a high intra- and inter-zone mobility, record the highest exposed count. Since there are few trips with Staten Island as destination, it has a low exposed count. As per the COVID-19 Tracking Project and the Center for Systems Science and Engineering at Johns Hopkins University, Queens and Brooklyn are truly the worst affected, as of November 2020 [52].

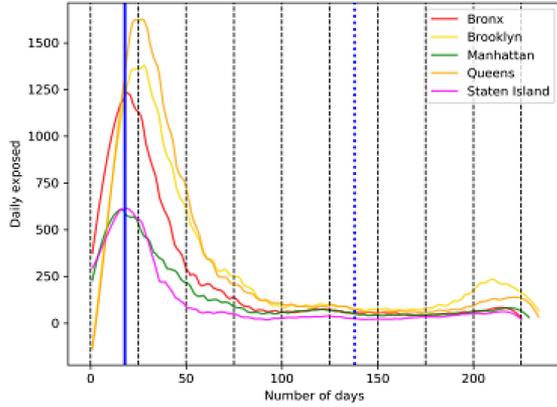


Fig. 10. Daily exposed curve for each borough and the starting and ending dates for lockdown is shown as a blue vertical line.

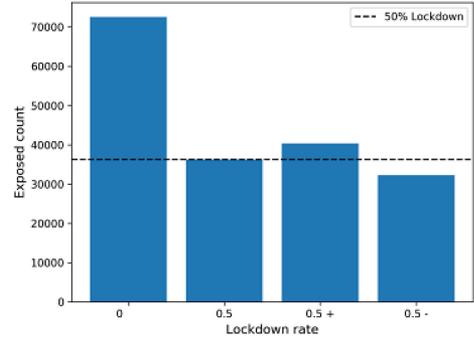


Fig. 12. The exposed numbers corresponding to four lockdown scenarios.

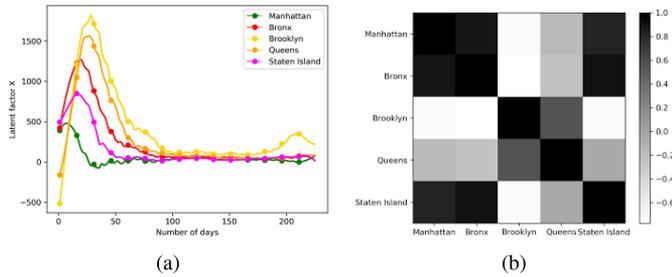


Fig. 11. Effect of Latent factor on infection spread and its variation during lockdown: (a) latent factor for each borough, (b) cosine similarity of mixing factors of NYC boroughs.

1) Peaking of the Exposed Curve and the Effect of Lockdown:

We plot the variation in daily exposed numbers in each borough (Fig. 10). Lockdown was formally initiated in the state of New York on March 20, 2020 [53], which is shown in solid blue line. Note that the exposed numbers briefly continued to rise for a week after the imposition of lockdown. However, the exposed curve is showing new peaks since October 2020. Finally, the exposed curves corresponding to Brooklyn and Queens peak at nearly the same time due to the high mobility between the two boroughs as depicted in Fig. 7(b).

2) *Latent Factor (X) Analysis:* We discuss in Section I that infection spread is not merely a function of human mobility, but a joint effect of mobility and social mixing, e.g., Bronx, despite its low inter-zone mobility has relatively high daily exposed numbers. We quantify the combination of mobility and mixing as a latent factor (X) (Section II-B). When we rank the boroughs in the non-increasing order of ICV of exposed, we see the following order: Manhattan (0.84), Brooklyn (0.80), Queens (0.70), Bronx (0.69) and Staten Island (0.63). In Fig. 11(a), we plot latent factor X_{it} for each borough. Observe that Queens and Brooklyn once again exhibit the highest x values. We already know that the latent factor is a combination of trip frequency, infected fraction and social mixing (see II-B3), we calculate the mixing factor (ι) from sampled X (using II-B3) for each borough. We apply mean-centered cosine similarity (Section II-C) to show (with

heatmap in Fig. 11(b)) that regions with high inter-zone mobility also show similar mixing, reinforcing infection spread.

3) *Lockdown Policymaking:* In Section II-B3, we discuss that the latent factor X can be scaled down by a fractional lockdown rate η , where $\eta = 1$ and 0 corresponds to no lockdown and complete lockdowns respectively. Using the new latent factor matrix $X' = \eta \times X$, we obtain a resultant exposed count $\Gamma' = A.X'$, where $\gamma'_{i,t} = \gamma_{i,t} \times \eta$ ($\forall \gamma_{i,t} \in \Gamma$). It is worth mentioning that the knowledge of the latent factor for each borough i at time t $x_{i,t}$ allows us to determine the ideal time and extent for η in order to minimize contagion. To prove our point, we introduce a vector of time varying eta at each timepoint t , $\eta = \{\eta_t | t = 1, 2, \dots, T\}$ and calculate X' by scaling the t -th column of X (denoted by $X_{*,t}$) by η_t .

Given X , let τ_{max} and τ_{min} (each of length $0 < k \leq T$) be two sets of timepoints with the highest and lowest sum of $X_{*,t}$, respectively. We consider the following scenarios: $\eta_t =$

- **Case 0:** $[1, 1, \dots, T$ times]
- **Case 0.5:** $[0.5, 0.5, \dots, T$ times]
- **Case 0.5+:** Same as Case 0.5, except overwrite η_t with 0.75 and 0.25 if $t \in \tau_{max}$ and $t \in \tau_{min}$, respectively.
- **Case 0.5-:** Same as Case 0.5, except overwrite η_t with 0.25 and 0.75 if $t \in \tau_{max}$ and $t \in \tau_{min}$, respectively.

We plot the total exposed in NYC in the pre-lockdown period for the four scenarios. Fig. 12 shows that we get the highest exposed for no lockdown (i.e., Case 0) and exactly half the exposed for 50% lockdown (Case 0.5). In Case 0.5+, we assign lesser lockdown ($\eta_t = 0.75$) to a timepoint with higher latent factor sum, and vice versa, and end up with a higher exposed sum than Case 0.5. Finally, in Case 0.5-, we achieve a considerably lower exposed by enforcing a higher lockdown ($\eta_t = 0.25$) at timepoints with higher latent factor sum. This suggests that the latent factors can help identify the ideal timepoints of imposing lockdowns to curb spread.

We apply principal component analysis (PCA) to visualize the reflection of the ensuing lockdown on the latent factor. The latent factor is a two-dimensional vector X comprising data-points $\{(x_{1t}, x_{2t}, \dots) | t = 1, 2, \dots, T\}$. Each point on the PCA plot (Fig. 13(a)) corresponds to a timepoint. We identify four clusters of timepoints, namely pre-lockdown (March 3 - 11), early-lockdown (March 22 - April 6), later-lockdown (April 21 - July 20) and post-lockdown (July 21 - October 15). In Fig. 13(b), we compare the daily infected numbers for these

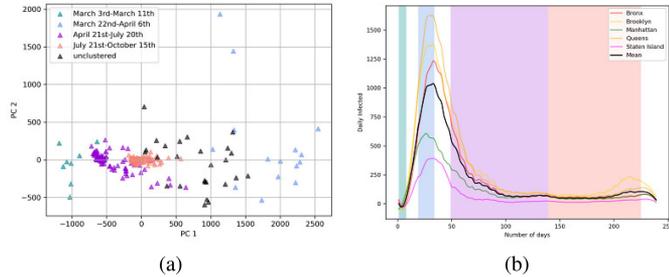


Fig. 13. Effect of lockdown. (a) principal component analysis (with two components) of the latent factor showing four clusters of time intervals, (b) comparison of the projected infected numbers for the four lockdown timelines.

timelines to show how the infection peaked from pre-lockdown to early lockdown phases and subsided thereafter.

D. Exploratory Analysis

The results presented so far show that the proposed spatiotemporal model is a generalized approach that can make informed spread predictions. It is worth highlighting how this model differs from existing efforts to study the evolution of contagion and the effects of public health intervention measures. Sun *et al.* adapt the susceptible, exposed, infected, confirmed and removed (comprising recovered and death) to model COVID-19 transmission at Wuhan, China. This model uses additional parameters to quantify infection coefficients under lockdown as well as emigration and immigration rates [54]. Similarly, Tian *et al.* performed curve-fitting on the time series of cases reported in Hubei province to learn the SEIR model parameters for COVID-19 and reported the immediate effect of lockdown on curbing the rate of contagion [55]. These models are instances of top-down approaches relying on applying curve-fitting to learn the SEIR model parameters. The proposed approach, however, is inherently different as it uses a simplified version of the SEIR model based on the daily infection counts alone, and in doing so, identifies a zone-specific measure of spread. Its benefit lies in the fact that it (1) reduces the number of parameters to be used in the SEIR model for fitting, (2) incorporates the variations in spread dynamics due to the exact inter- and intra-zone mobility patterns, and (3) lends itself to a more generalized time-varying analysis that a traditional fitting-based approach may fail to capture.

Let us discuss how this model inspires several research directions to design mobility policies to combat future outbreaks. First, this model can infer the effect of trip lengths on the exposed numbers. As discussed in Section II-B, we decompose the latent factor X into trip frequency, infected ratio and mixing factor. For instance, the exposed population count as a result of trips of length $r + 1$ or more can be expressed as $E_{\geq r} = \hat{A}_{>r} \times X$; here $A_{>r} = \sum_{i=r+1}^{\zeta} A^i$ (refer (6)) are the stochastic matrices estimating probability of trips of length $r + 1$ or more. One may approximate $A_{>r}$ to $\sum_{i=r+1}^{\zeta} A^i$, where $|\sum_{i=r+1}^{\zeta+1} A^i - \sum_{i=r+1}^{\zeta} A^i|$ is less than some threshold th . Given that the average inter- and intra- borough distance in NYC is 14.4 miles, for $r = 12$, we observe that $E_{>12}$ accounts for a little over 1% of the total exposed numbers. It is worth

noting that since the length of 65% of trips in NYC are less than 2 miles, the pandemic spread can be contained effectively by restricting shorter trips, as longer trips have little bearing on the exposed numbers. Second, the exposed numbers often lends great insights into the time it would take to flatten the curve. Nonlinear curve-fitting on the daily exposed numbers for NYC boroughs explain why (in the current state of lockdown) the new exposed numbers started dropping by the end of May and daily infected curve (that lagged by roughly a week) started stabilizing by early June.

IV. CONCLUSION

COVID-19 has insidiously affected every facet of human existence over the last 11 months. Global infected and death numbers for COVID-19 suggest that geography, time, and mobility patterns are key factors affecting spread, making it imperative to factor in the spatiotemporal and mobility context in future prediction models. In this work, we present a spatiotemporal model for pandemic spread that unifies mobility and social mixing into a latent factor. We apply the model on the NYC data to show that boroughs of high inter-zone mobility (namely, Brooklyn, and Queens exhibit similar trends in exposed numbers as well as mixing. We carry out principal component analysis to depict the temporal variation of the latent factor in pre- and post-lockdown epochs. Next, we argue that the inverse coefficient of variation (ICV) of daily exposed curve can explain the duration as well as the spread of infection in a zone. We show that the ICV of nations (such as USA, Iran, China, New Zealand, etc.) correlate with the true extent and period of COVID-19 spread. We show the validity of the proposed method by estimating the exposed numbers for different countries. Furthermore, we report that the peaks in daily exposed of neighboring nations are reached at nearly the same time, underpinning the role of proximity in spread.

The discussion on lockdown policies (Section III-C3) and exploratory analysis (Section III-D) shows that the proposed spatiotemporal model motivates new research directions for studying the mixing propensity of individuals at both inter- and intra-borough levels. The drop in the latent factor (X) post-lockdown was achieved by a reduced number of trips (although we will consider real post-lockdown mobility traces in our future experiments); however, lockdown or social distancing measures are expected to impact the mixing propensity of individuals in different contexts, such as mixing at grocery stores, restaurants, work-places or at home. Ideally, the mixing factor can be considered as a vector of latent items each of which may be impacted differently by constraints on individual mobility. These observations can feed into future SEIR models and quarantine procedures to limit the infection spread. Since regions with high inter-zone mobility patterns exhibit similar trends in exposed numbers, it is advisable to consider clusters of such zones in future infection models. A SEIR model executing on a single zone needs to otherwise consider the high inflow/outflow rates of susceptible/infected individuals into and out of the zones which again increases model complexity and can make the parameters non-identifiable and hence hard to justify [16]. A clustered view of zones with high mobility and mixing keeps the

models simpler to the general SEIR format and helps generate more realistic predictions. Similarly, quarantine procedures on such entire clusters can help better contain the infection spread while additionally alleviating the economic loss that results from limiting the two-hop trips.

REFERENCES

- [1] "Coronavirus: What have been the worst pandemics and epidemics in history?" 2020. [Online]. Available: https://en.as.com/en/2020/04/18/other_sports/1587167182_422066.html
- [2] "Coronavirus world map: Which countries have the most cases and deaths?" 2020. [Online]. Available: <https://www.theguardian.com/world/2021/jan/28/covid-world-map-which-countries-have-the-most-coronavirus-vaccinations-cases-and-deaths>
- [3] S. 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," *Infect. Dis. Poverty*, vol. 9, no. 1, p. 29, 2020, Art. no. 29.
- [4] N. Khan, M. Naushad, S. Fahad, S. Faisal, and A. Muhammad, "Covid-2019 and world economy," *J. Health Econ., Forthcoming*, 2020, doi: [10.2139/ssrn.3566632](https://doi.org/10.2139/ssrn.3566632).
- [5] S. R. Baker, N. Bloom, S. J. Davis, and S. J. Terry, "Covid-induced economic uncertainty," *Nat. Bur. Econ. Res.*, Work. Paper 26983, Apr. 2020, doi: [10.3386/w26983](https://doi.org/10.3386/w26983).
- [6] "709000 seek US unemployment aid as covid-19 pandemic escalates," 2020. [Online]. Available: <https://www.usatoday.com/story/money/2020/11/12/unemployment-709000-seek-us-jobless-aid-covid-19-cases-spike/6263936002/>
- [7] R. Anderson, H. Heesterbeek, D. Klinkenberg, and T. Hollingsworth, "How will country-based mitigation measures influence the course of the covid-19 epidemic?," *The Lancet*, vol. 395, no. 10228, pp. 931–934, 2020.
- [8] "A fiasco in the making? as the coronavirus pandemic takes hold, we are making decisions without reliable data," 2020. [Online]. Available: <https://gwire.com/2020/03/21/a-fiasco-in-the-making-as-the-coronavirus-pandemic-takes-hold-we-are-making-decisions-without-reliable-data/>
- [9] "10 reasons to doubt the covid-19 data," 2020. [Online]. Available: <https://www.bloomberg.com/opinion/articles/2020-04-13/ten-reasons-to-doubt-the-covid-19-data>
- [10] "Coronavirus: It's time to get real about the misleading data," 2020. [Online]. Available: <https://thehill.com/opinion/technology/490541-coronavirus-its-time-to-get-real-about-the-misleading-data>
- [11] R. Niehus, P. De Salazar, A. Taylor, and M. Lipsitch, "Using observational data to quantify bias of traveller-derived covid-19 prevalence estimates in Wuhan, China," *The Lancet Infect. Dis.*, vol. 20, no. 7, pp. 803–808, 2020.
- [12] Z. Yang *et al.*, "Modified SEIR and AI prediction of the epidemics trend of Covid-19 in China under public health interventions," *J. Thoracic Dis.*, vol. 12, no. 3, pp. 165–174, 2020.
- [13] Y. Fang, Y. Nie, and M. Penny, "Transmission dynamics of the covid-19 outbreak and effectiveness of government interventions: A data-driven analysis," *J. Med. Virol.*, vol. 92, no. 6, pp. 645–659, 2020.
- [14] H. Hethcote, "The mathematics of infectious diseases," *SIAM Rev.*, vol. 42, no. 4, pp. 599–653, 2000.
- [15] N. P. Jewell, J. A. Lewnard, and B. L. Jewell, "Caution warranted: Using the institute for health metrics and evaluation model for predicting the course of the covid-19 pandemic," *Ann. Intern. Med.*, vol. 173, no. 3, pp. 226–227, 2020.
- [16] W. Roda, M. Varughese, D. Han, and M. Li, "Why is it difficult to accurately predict the covid-19 epidemic?," *Infect. Dis. Modelling*, vol. 5, pp. 271–281, 2020.
- [17] M. Kraemer *et al.*, "The effect of human mobility and control measures on the covid-19 epidemic in china," *Science*, vol. 368, no. 6490, pp. 493–497, 2020.
- [18] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of Covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, 2020, doi: [10.1136/bmj.m1328](https://doi.org/10.1136/bmj.m1328).
- [19] A. Alimadadi, S. Aryal, I. Manandhar, P. Munroe, B. Joe, and X. Cheng, "Artificial intelligence and machine learning to fight covid-19," *Physiol. Genomic.*, vol. 52, no. 4, pp. 200–202, 2020.
- [20] G. Randhawa, M. Soltysiak, H. El Roz, C. de Souza, K. Hill, and L. Kari, "Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: Covid-19 case study," *Plos One*, vol. 15, no. 4, 2020, Art. no. e0232391.
- [21] I. Holmdahl and C. Buckee, "Wrong but useful-what covid-19 epidemiologic models can and cannot tell us," *New England J. Med.*, vol. 383, no. 4, pp. 303–305, 2020.
- [22] P. Wang, X. Zheng, J. Li, and B. Zhu, "Prediction of epidemic trends in covid-19 with logistic model and machine learning technics," *Chaos, Solitons Fractals*, vol. 139, 2020, Art. no. 110058.
- [23] S. Roy and P. Ghosh, "Factors affecting covid-19 infected and death rates inform lockdown-related policymaking," *PLoS One*, vol. 15, no. 10, 2020, Art. no. e0241165.
- [24] S. He, Y. Peng, and K. Sun, "Seir modeling of the covid-19 and its dynamics," *Nonlinear Dyn.*, vol. 101, no. 3, pp. 1667–1680, 2020.
- [25] G. Pandey, P. Chaudhary, R. Gupta, and S. Pal, "Seir and regression model based covid-19 outbreak predictions in India," 2020, available: [arXiv:2004.00958](https://arxiv.org/abs/2004.00958).
- [26] S. Annas *et al.*, "Stability analysis and numerical simulation of seir model for pandemic covid-19 spread in indonesia," *Chaos, Solitons Fractals*, vol. 139, 2020, Art. no. 110072.
- [27] A. Radulescu and K. Cavanagh, "Management strategies in a seir model of covid 19 community spread," 2020, available: [arXiv:2003.11150](https://arxiv.org/abs/2003.11150).
- [28] K. Iwata and C. Miyakoshi, "A simulation on potential secondary spread of novel coronavirus in an exported country using a stochastic epidemic seir model," *J. Clin. Med.*, vol. 9, no. 4, 2020, Art. no. 944, doi: [10.3390/jcm9040944](https://doi.org/10.3390/jcm9040944).
- [29] S. Mwalili, M. Kimathi, V. Ojiambo, D. Gathungu, and R. Mbogo, "Seir model for covid-19 dynamics incorporating the environment and social distancing," *BMC Res. Notes*, vol. 13, 2020, Art. no. 352.
- [30] Z. Tang, X. Li, and H. Li, "Prediction of new coronavirus infection based on a modified seir model," *medRxiv*, 2020, doi: [10.1101/2020.03.03.20030858](https://doi.org/10.1101/2020.03.03.20030858).
- [31] L. López and R. Xavier, "A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: Simulating control scenarios and multi-scale epidemics," *Res. Phys.*, vol. 21, 2021, Art. no. 103746.
- [32] H. Abdi, "Coefficient of variation," *Encyclopedia Res. Des.*, vol. 1, pp. 169–171, 2010.
- [33] W. Sharpe, "The sharpe ratio," *J. Portfolio Manage.*, vol. 21, no. 1, pp. 49–58, 1994.
- [34] B. Nordstrom, "Finite markov chains," 2008, <http://www.math.uchicago.edu/may/VIGRE/VIGRE2008/REUPapers/Nordstrom.pdf>
- [35] J. Henseler and G. Fassott, "Testing moderating effects in pls path models: An illustration of available procedures," *Handbook Partial Least Squares*. Berlin, Germany: Springer, 2010, pp. 713–735, doi: [10.1007/978-3-540-32827-8_31](https://doi.org/10.1007/978-3-540-32827-8_31).
- [36] X. Ning, C. Desrosiers, and G. Karypis, "A comprehensive survey of neighborhood-based recommendation methods," *Recommender Systems Handbook*. Berlin, Germany: Springer, 2015, pp. 37–76, doi: [10.1007/978-1-4899-7637-6_2](https://doi.org/10.1007/978-1-4899-7637-6_2).
- [37] "Coronavirus incubation period," 2020. [Online]. Available: <https://www.news-medical.net/health/Coronavirus-Incubation-Period.aspx>
- [38] "scientists revise covid-19 incubation period to 7.7 days," 2020. [Online]. Available: <https://www.medicalnewstoday.com/articles/scientists-revise-covid-19-incubation-period-to-7-7-days#Longer-incubation-period>
- [39] "Nyc - the Official Website of New York City," 2020. [Online]. Available: <https://www1.nyc.gov/site/doh/data/data-publications/profiles.page>
- [40] "Geopy: Geocoding library for python," 2020. [Online]. Available: <https://github.com/geopy/geopy>
- [41] Nycopendata, 2020. [Online]. Available: <https://data.cityofnewyork.us/Transportation/Traffic-Volume-Counts-2012-2013-/p424-amsu>
- [42] "Nyc department of health and mental hygiene," 2020. [Online]. Available: <https://github.com/nychealth/coronavirus-data>
- [43] W. H. Organization, "Humanitarian data exchange covid-19 dataset," 2020. [Online]. Available: <https://data.humdata.org/dataset/coronavirus-covid-19-cases-and-deaths/resource/2ac6c3c0-76fa-4486-9ad0-9aa9e253b78d>
- [44] T. Burki, "China's successful control of covid-19," *The Lancet Infect. Dis.*, vol. 20, no. 11, pp. 1240–1241, 2020.
- [45] "How did New Zealand control covid-19?" 2020. [Online]. Available: <https://www.contagionlive.com/view/how-did-new-zealand-control-covid19>
- [46] "Physician predicts spike in US covid-19 cases after thanksgiving," 2020. [Online]. Available: <https://www.cnn.com/2020/11/14/health/us-coronavirus-saturday/index.html>
- [47] "'catastrophic': Iran reports record rise in covid-19 cases," 2020. [Online]. Available: <https://www.aljazeera.com/news/2020/10/6/catastrophic-iran-reports-record-rise-in-covid-19-cases>

- [48] K. Lee, S. Hong, S. Kim, I. Rhee, and S. Chong, "Slaw: Self-similar least-action human walk," *IEEE/ACM Trans. Netw.*, vol. 20, no. 2, pp. 515–529, Apr. 2012.
- [49] J. Ghosh, S. Philip, and C. Qiao, "Sociological orbit aware location approximation and routing (solar) in manet," *Ad Hoc Netw.*, vol. 5, no. 2, pp. 189–209, 2007.
- [50] M. Musolesi and C. Mascolo, "Designing mobility models based on social network theory," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 11, no. 3, pp. 59–70, 2007.
- [51] R. Storn and K. Price, "Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces," *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, 1997.
- [52] "Coronavirus in New York city," 2020. [Online]. Available: https://projects.thecity.nyc/2020_03_covid-19-tracker/
- [53] "Coronavirus in Ny: Cuomo orders lockdown, shuts down non-essential businesses," 2020. [Online]. Available: <https://nypost.com/2020/03/20/coronavirus-in-ny-cuomo-orders-lockdown-shuts-down-non-essential-businesses/>
- [54] G. Sun *et al.*, "Transmission dynamics of covid-19 in wuhan, china: Effects of lockdown and medical resources," *Nonlinear Dyn.*, vol. 101, no. 3, pp. 1981–1993, 2020.
- [55] H. Tian *et al.*, "An investigation of transmission control measures during the first 50 days of the covid-19 epidemic in china," *Science*, vol. 368, no. 6491, pp. 638–642, 2020.



Satyaki Roy received the Ph.D. degree in computer science from the Missouri University of Science and Technology, Rolla, MO, USA, in 2019. He is currently a Postdoctoral Research Associate with the Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. His research interests include computational biology, network science and optimization, wireless sensor networks, epidemiology, machine learning, and parallel computing.



Preetom Biswas is currently an undergraduate student with the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh. He is currently a Math and Programming Enthusiast. His research interests include the application of machine learning and graph theoretic approaches to solving real-world problems.



Preetam Ghosh received the B.S. degree in computer science from Jadavpur University, Kolkata, India, and the M.S. and Ph.D. degrees in computer science and engineering from the University of Texas at Arlington, Arlington, TX, USA. He is currently a Professor with the Department of Computer Science and directs the Biological Networks Lab, Virginia Commonwealth University, Richmond, VA, USA. His research interests include algorithms, stochastic modeling and simulation, network science and machine learning related approaches in systems biology and compu-

tational epidemiology and mobile computing related issues in pervasive grids that have resulted in more than 170 conferences and journal articles and several federally funded research projects from the NSF, the NIH, the DoD and the US-VHA. He is currently the Secretary or the Treasurer of the ACM SIGBio.