

Analysis and Applications of Population Flows in a Networked SEIRS Epidemic Process

Brooks A. Butler, Raphael Stern, and Philip E. Paré*

Abstract—Transportation networks play a critical part in the spread of infectious diseases between populations. In this work, we define a networked susceptible-exposed-infected-recovered epidemic process with loss of immunity over time (SEIRS) that explicitly models the flow of individuals between sub-populations, which serves as the propagating mechanism for infection. We provide sufficient conditions for local stability and instability of the healthy state of the system and show that no perturbation of population flows can change the local stability of any healthy state. We also provide sufficient conditions for the existence and uniqueness of an endemic state. We then develop tools and methods for applying our model to real-world data, including spreading parameter estimation and disease arrival time prediction, and apply them in a case study using both travel and infection data from counties in Minnesota during the first year of the COVID-19 pandemic.

Index Terms—Epidemic Spread, Transportation, Network Modeling.

1 INTRODUCTION

With the prevalence of transportation networks that enable rapid travel between locations globally, understanding how these transportation networks facilitate the spread of infectious diseases only continues to grow in importance. While certainly not the first globally transmitted disease, the recent COVID-19 pandemic has demonstrated the negative effect that a highly infectious disease can have on general health and safety of any population connected via travel [1]–[6], as well as long-lasting impacts on both local and global economies [7]–[10]. Given the growing impact and scale of human mobility as global interconnectivity continues to increase, as well as the near certainty of other pandemic-level diseases emerging in the future, it is critical that we develop a rigorous and sophisticated understanding of such phenomena in the networked sense. Thus, in this work, we aim to contribute to a deeper understanding of how transportation networks may be used to model and predict the transmission of infectious diseases over a network of connected populations.

This work builds off the deep body of literature dedicated to modeling epidemic processes using compartmental models [11]–[15]. Such models separate a population, or individuals in a population, into distinct compartments

which represent their infection state. Simple compartmental models may have as few as two states, such as being either susceptible to a given disease or infected. For diseases with more complex behavior, more compartments may be needed to better describe the progression of a population through the epidemic process of the disease. In this work, we focus on a four-compartment model which separates individuals into one of four states, namely: susceptible, exposed, infected, and recovered. Further, we allow for a rate at which a population or individual may lose immunity to the given disease, and therefore may become susceptible again. This model, abbreviated to the SEIRS model, allows for the modeling of a disease that has a period in which an individual may be exposed and will become infectious, but is not immediately apparent. Additionally, the recovered state and eventual loss of immunity model a disease that can evolve over time, but may allow for temporary immunity after an initial recovery. This model is chosen, in part, to better emulate what is now known of the behavior of COVID-19 and its variants [16]–[19]. Compartmental models alone, however, do not account for the transmission of a disease in a networked environment where a general population can be divided into subpopulations connected via various transportation networks [20]–[29]. Thus, an increase in scope is needed to account for this important behavioral complexity in disease transmission.

Previous work that incorporates networked population flows in an epidemic process model includes analysis of both networked SIS and SIR models with flows [30], [31]. Other work uses travel flows to compute the effective distance between population centers to predict arrival times for various epidemics using global and local travel data [32]–[36]. Additionally, a data-driven approach to modeling the effect of population flow on epidemic spreading is performed in [37], which uses data from lockdowns in Italy during the initial spreading of COVID-19. Additional work on epidemic spreading with population flow has been performed in modeling multi-city epidemics [38]–[41] and in the modeling of disease spread in metapopulations [42]–[45]. We contribute to the development of such networked models by including the exposed state in our model formulation as well as a loss of immunity. In our previous work, we consider capturing the effect of transportation on the spread of COVID-19 using a networked discrete-time SEIR model [46], [47]; however, the key distinction in this work is that, in addition to being a continuous-time model, infection propagation over the network is modeled by the relocation of infected individuals to other

*Brooks A. Butler is with the Department of Electrical Engineering and Computer Science at the University of California, Irvine, Email: {bbutler2@uci.edu}. Raphael Stern is with the Department of Civil, Environmental, and Geo-Engineering at the University of Minnesota, Email: {rstern@umn.edu}. Philip E. Paré is with the Elmore Family School of Electrical and Computer Engineering at Purdue University, Email: {phil-paré@purdue.edu}. This work was funded in part by the National Science Foundation, grants NSF-CNS #2028946 (R.S.), and NSF-CNS #2028738, NSF-ECCS #2032258 and #2238388 (P.E.P.).

sub-populations rather than assuming direct contact and intermingling between sub-populations. Further, we allow recovered individuals to experience a loss of immunity over time, which creates the possibility of an endemic state for the epidemic process.

In this paper, we make the following contributions:

- 1) We define a networked SEIRS model that incorporates travel flows as a means of disease propagation.
- 2) We provide conditions for local stability and instability of the healthy state and prove that perturbations to population flows cannot change the stability of the healthy state
- 3) We provide conditions under which a unique endemic state will exist.
- 4) We implement methods for estimating the spreading parameters of our model using both infection and travel data as well as a method for predicting the arrival time of a disease in the case of multiple outbreak locations using travel data to compute the effective distance between uninfected nodes and the infected group.
- 5) We demonstrate how these tools can be applied to our model using real-world mobility data from Minnesota during the COVID-19 pandemic.

We now provide the necessary definitions of notation and previous analytical results used in our model construction and analysis sections, respectively.

2 PRELIMINARIES

We denote the set of real numbers and positive real numbers as $\mathbb{R}$ and $\mathbb{R}_{>0}$, respectively. For any positive integer n , we have $[n] = \{1, 2, \dots, n\}$. A diagonal matrix is denoted as $\text{diag}(\cdot)$. The transpose of a vector $x \in \mathbb{R}^n$ is $x^\top$. We use $\mathbf{0}$ and $\mathbf{1}$ to denote the vectors of the appropriate size whose entries all equal 0 and 1, respectively. We let $\mathcal{G} = (\mathbb{V}, \mathbb{E}, \mathbb{W})$ denote a weighted directed graph where $\mathbb{V} = \{v_1, v_2, \dots, v_n\}$ is the set of nodes, $\mathbb{E} \subseteq \mathbb{V} \times \mathbb{V}$ is the set of edges, and $\mathbb{W} : \mathbb{E} \rightarrow \mathbb{R}_{>0}$ maps to the real valued edge weights on each edge. We denote the configuration of edges in a directed graph at time t as $\mathcal{G}(t) = (\mathbb{V}, \mathbb{E}(t), \mathbb{W})$, where $\mathbb{E}(t)$ denotes the set of edges at time t . Furthermore, we denote $\cup_{t \geq 0} \mathbb{E}(t)$ as the union of all non-zero edge configurations on a graph for all $t \geq 0$. We define a graph $\mathcal{G}$ as being strongly connected if there is a path consisting of nonzero edge weights from every node to every other node in the graph.

For a complex number x we use $|x|$ and $\text{Re}(x)$ to denote its magnitude and real part, respectively. For a real square matrix M , we use $\rho(M)$ to denote its spectral radius and $s(M)$ to denote the largest real part amongst its eigenvalues, i.e.,

$$\begin{aligned} \rho(M) &= \max\{|\lambda| : \lambda \in \sigma(M)\}, \\ s(M) &= \max\{\text{Re}(\lambda) : \lambda \in \sigma(M)\}, \end{aligned}$$

where $\sigma(M)$ denotes the spectrum of M .

A real square matrix is called *Metzler* if its off-diagonal entries are all non-negative. Thus, any non-negative matrix is Metzler.

Lemma 1. For any matrix M and any real number ϕ , if $A := M - \phi I$, then $\sigma(M) = \sigma(A) + \phi$.

The following results from Chapter 2 of [48] for non-negative matrices, which also hold for Metzler matrices by Lemma 1, with $\phi = \min\{0, m_{11}, \dots, m_{nn}\}$, will be used in the subsequent analysis.

Lemma 2. (Lemma 2.3 in [48]) Suppose that M is an irreducible Metzler matrix. Then, $s(M)$ is a simple eigenvalue of M and there exists a unique (up to scalar multiple) vector $x \gg \mathbf{0}$ such that $Mx = s(M)x$.

Proposition 1. Suppose that Λ is a negative diagonal matrix in $\mathbb{R}^{n \times n}$ and N is an irreducible non-negative matrix in $\mathbb{R}^{n \times n}$. Let $M = \Lambda + N$. Then, $s(M) < 0$ if and only if $\rho(-\Lambda^{-1}N) < 1$, $s(M) = 0$ if and only if $\rho(-\Lambda^{-1}N) = 1$, and $s(M) > 0$ if and only if $\rho(-\Lambda^{-1}N) > 1$.

Lemma 3. (Levy-Desplanques Theorem) A strictly diagonally dominant matrix is non-singular. In other words, let $A \in \mathbb{C}^{n \times n}$ be a matrix satisfying the property

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|, \quad \forall i;$$

then $\det(A) \neq 0$.

Lemma 4. (Observation 6.3.1 in [49]) Let $M \in \mathbb{R}^{n \times n}$ be diagonalizable with $M = S\Lambda S^{-1}$ and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ and let $E \in \mathbb{R}^{n \times n}$. If $\hat{\lambda}$ is an eigenvalue of $M + E$, then there exists some eigenvalue λ_i of M for which

$$|\hat{\lambda} - \lambda_i| \leq \|S\|_\infty \|S^{-1}\|_\infty \|E\|_\infty = \kappa_\infty(S) \|E\|_\infty,$$

where $\kappa_\infty(S)$ denotes the condition number with respect to the infinity matrix norm $\|\cdot\|_\infty$.

3 NETWORK FLOWS MODEL DEFINITION

We now present a networked SEIRS model incorporating the population flow of individuals between subpopulations. First, consider a group of n sub-populations in a graph, where each sub-population $i \in [n]$ is represented by a node in the graph $\mathcal{G}$. We use the SEIRS model to describe how susceptible individuals in sub-population i become exposed, infected, recover, and gradually lose immunity as the result of an infectious disease [50]. We begin with defining the SEIRS model behavior without graph connections for each sub-population $i \in [n]$. Let S_i , E_i , I_i , and R_i represent the number of susceptible, exposed, infected, and recovered individuals in sub-population i , respectively, and their dynamics evolve as

$$\dot{S}_i(t) = \alpha_i R_i(t) - \beta_i \frac{I_i(t)}{N_i} S_i(t) \quad (1a)$$

$$\dot{E}_i(t) = \beta_i \frac{I_i(t)}{N_i} S_i(t) - \sigma_i E_i(t) \quad (1b)$$

$$\dot{I}_i(t) = \sigma_i E_i(t) - \delta_i I_i(t) \quad (1c)$$

$$\dot{R}_i(t) = \delta_i I_i(t) - \alpha_i R_i(t), \quad (1d)$$

where β_i is the infection rate, σ_i is the transition rate from exposed to infected, δ_i is the healing rate, and α_i is the

rate of immunity loss. Although the state variables, except population, will continue to vary with time, we remove the time-dependence notation for convenience and ease of reading from this point forward.

To account for the flow of individuals between sub-populations we expand the model in (1):

$$\dot{S}_i = -\beta_i \frac{I_i}{N_i} S_i + \alpha_i R_i + \sum_{j \neq i} \left(F_{ij} \frac{S_j}{N_j} - F_{ji} \frac{S_i}{N_i} \right) \quad (2a)$$

$$\dot{E}_i = \beta_i \frac{I_i}{N_i} S_i - \sigma_i E_i + \sum_{j \neq i} \left(F_{ij} \frac{E_j}{N_j} - F_{ji} \frac{E_i}{N_i} \right) \quad (2b)$$

$$\dot{I}_i = \sigma_i E_i - \delta_i I_i + \sum_{j \neq i} \left(F_{ij} \frac{I_j}{N_j} - F_{ji} \frac{I_i}{N_i} \right) \quad (2c)$$

$$\dot{R}_i = \delta_i I_i - \alpha_i R_i + \sum_{j \neq i} \left(F_{ij} \frac{R_j}{N_j} - F_{ji} \frac{R_i}{N_i} \right), \quad (2d)$$

where F_{ij} represents the number of individuals flowing from sub-population j to i , with $F_{ii} = 0$. By making a substitution of variables where $s_i = S_i/N_i$, $e_i = E_i/N_i$, $x_i = I_i/N_i$, $r_i = R_i/N_i$ we can model the proportion of individuals as follows

$$\dot{s}_i = \alpha_i r_i - \beta_i x_i s_i + \frac{1}{N_i} \sum_{j \neq i} (F_{ij} s_j - F_{ji} s_i) \quad (3a)$$

$$\dot{e}_i = \beta_i x_i s_i - \sigma_i e_i + \frac{1}{N_i} \sum_{j \neq i} (F_{ij} e_j - F_{ji} e_i) \quad (3b)$$

$$\dot{x}_i = \sigma_i e_i - \delta_i x_i + \frac{1}{N_i} \sum_{j \neq i} (F_{ij} x_j - F_{ji} x_i) \quad (3c)$$

$$\dot{r}_i = \delta_i x_i - \alpha_i r_i + \frac{1}{N_i} \sum_{j \neq i} (F_{ij} r_j - F_{ji} r_i), \quad (3d)$$

where $s_i + e_i + x_i + r_i = 1$. Note that both (2) and (3) assume the subpopulations are well mixed and that the likelihood of an individual traveling is independent of their infectious state, that is, whether they are susceptible, exposed, infected, or recovered. We can compute the number of individuals flowing from sub-population j to i as $F_{ij} = \gamma_j w_{ij} N_j$, where γ_j is the proportion of the total population flowing out of node j computed as $\gamma_j = \frac{\sum_{i \neq j} F_{ij}}{N_j}$, and w_{ij} is the proportion of traveling individuals flowing from sub-population j to i computed as $w_{ij} = \frac{F_{ij}}{\sum_{i \neq j} F_{ij}}$, with $w_{ii} = 0$. Thus, we can further derive the dynamics for the susceptible proportion at sub-population i as

$$\begin{aligned} \dot{s}_i &= \alpha_i r_i - \beta_i x_i s_i + \frac{1}{N_i} \sum_{j \neq i} (F_{ij} s_j - F_{ji} s_i) \\ &= \alpha_i r_i - \beta_i x_i s_i + \frac{1}{N_i} \sum_{j \neq i} (\gamma_j w_{ij} N_j s_j - \gamma_i w_{ji} N_i s_i) \\ &= \alpha_i r_i - \beta_i x_i s_i + \sum_{j \neq i} \left(\frac{N_j}{N_i} w_{ij} \gamma_j s_j - w_{ji} \gamma_i s_i \right). \end{aligned}$$

Using the fact that $\sum_{j \neq i} w_{ji} = 1$, we have that

$$\dot{s}_i = \alpha_i r_i - (\beta_i x_i + \gamma_i) s_i + \sum_{j \neq i} \frac{N_j}{N_i} w_{ij} \gamma_j s_j.$$

By similar derivations we can rewrite (3) as

$$\dot{s}_i = \alpha_i r_i - (\beta_i x_i + \gamma_i) s_i + \sum_{j \neq i} \frac{N_j}{N_i} w_{ij} \gamma_j s_j \quad (4a)$$

$$\dot{e}_i = \beta_i x_i s_i - (\sigma_i + \gamma_i) e_i + \sum_{j \neq i} \frac{N_j}{N_i} w_{ij} \gamma_j e_j \quad (4b)$$

$$\dot{x}_i = \sigma_i e_i - (\delta_i + \gamma_i) x_i + \sum_{j \neq i} \frac{N_j}{N_i} w_{ij} \gamma_j x_j \quad (4c)$$

$$\dot{r}_i = \delta_i x_i - (\gamma_i + \alpha_i) r_i + \sum_{j \neq i} \frac{N_j}{N_i} w_{ij} \gamma_j r_j. \quad (4d)$$

Note that our model does not allow for state transitions to occur during travel, i.e., we do not model individuals becoming infected as a result of their travel. Rather, individuals will change their state in the epidemic process at the rate defined at each node. Of course, this may not always be the case in reality, as some methods of public transportation may facilitate infection spreading for those sharing a confined space with inadequate ventilation. Therefore, this study implicitly focuses on more long-range modes of transportation that transfer individuals between more distinctly separated populations (such as travel via flights or long-range commuting) where traveling individuals may not be interacting with other travelers, or where there is a sufficient amount of ventilation such that the spread of the disease is significantly reduced during transit [51]. The inclusion of infections occurring on the edges of transportation networks is left as a direction for future work.

For the model in (4) to be well-defined, we required the following assumptions.

Assumption 1. Let $\sum_{i \neq j} F_{ji} = \sum_{i \neq j} F_{ij}$ for all $i, j \in [n]$.

This assumption requires that the total flow of individuals into a given sub-population must be equal to the total flow out, which naturally follows from the assumption that the population at each node remains constant. Furthermore, we impose the following assumption on the model parameters and initial conditions.

Assumption 2. Let $\alpha_i, \beta_i, \sigma_i, \delta_i \in \mathbb{R}_{>0}$, $s_i(t_0), e_i(t_0), x_i(t_0), r_i(t_0) \in [0, 1]$, and $s_i(t_0) + e_i(t_0) + x_i(t_0) + r_i(t_0) = 1$ for all $i \in [n]$ and $t_0 \in \mathbb{R}_{\geq 0}$.

Under these assumptions, we can show that the model will always remain well-defined.

Lemma 5. Let Assumptions 1-2 hold, then $s_i(t), e_i(t), x_i(t), r_i(t) \in [0, 1]$, and $s_i(t) + e_i(t) + x_i(t) + r_i(t) = 1$ for all $i \in [n]$ and for all $t \in \mathbb{R}_{\geq 0}$.

Proof. First, by taking the sum of the system dynamics as defined in (3) we have

$$\dot{s}_i(t) + \dot{e}_i(t) + \dot{x}_i(t) + \dot{r}_i(t) = 0, \quad (5)$$

by Assumption 1 at every time t . Thus, by integrating (5) from t_0 to t we have

$$\int_{t_0}^t \dot{s}_i(\tau) + \dot{e}_i(\tau) + \dot{x}_i(\tau) + \dot{r}_i(\tau) d\tau = 0, \\ \implies s_i(t) + e_i(t) + x_i(t) + r_i(t) = 1, t \in \mathbb{R}_{\geq 0}$$

by Assumption 2. To show that $s_i(t), e_i(t), x_i(t), r_i(t) \in [0, 1]$, we consider the system as each state approaches zero. Note that $\lim_{q_i(t) \rightarrow 0} \dot{q}_i(t) \geq 0$ where $q_i \in \{s_i, e_i, x_i, r_i\}$. Thus, by the assumption that the initial states are non-negative, we have that $q_i(t) \geq 0, \forall t \geq 0$. Further, since we have shown that $s_i(t) + e_i(t) + x_i(t) + r_i(t) = 1$ it follows directly that $s_i(t), e_i(t), x_i(t), r_i(t) \in [0, 1], \forall t \geq 0$ and $\forall i \in [n]$. $\square$

For the purpose of the forthcoming analysis, we also express the system in (4) in matrix form. Let $\Phi = N^{-1}WTN$, where $N = \text{diag}(N_i)$, $W \in \mathbb{R}^{n \times n}$ is the matrix comprised of the entries w_{ij} , and $\Gamma = \text{diag}(\gamma_i)$. The vectorized equations then become

$$\dot{s} = Ar - (BX(t) + \Gamma)s + \Phi s \quad (6a)$$

$$\dot{e} = BX(t)s - (\Sigma + \Gamma)e + \Phi e \quad (6b)$$

$$\dot{x} = \Sigma e - (D + \Gamma)x + \Phi x \quad (6c)$$

$$\dot{r} = Dx - (A + \Gamma)r + \Phi r, \quad (6d)$$

where $A = \text{diag}(\alpha_i)$, $B = \text{diag}(\beta_i)$, $\Sigma = \text{diag}(\sigma_i)$, $D = \text{diag}(\delta_i)$, $X(t) = \text{diag}(x_i)$, $s = [s_1, \dots, s_n]^\top$, $e = [e_1, \dots, e_n]^\top$, $x = [x_1, \dots, x_n]^\top$, and $r = [r_1, \dots, r_n]^\top$. We further construct the matrix

$$\underbrace{\begin{bmatrix} -BX(t) - \Gamma + \Phi & 0 & 0 & A \\ BX(t) & -\Sigma - \Gamma + \Phi & 0 & 0 \\ 0 & \Sigma & -D - \Gamma + \Phi & 0 \\ 0 & 0 & D & -A - \Gamma + \Phi \end{bmatrix}}_{H(t)}, \quad (7)$$

where

$$H(t) = -Q(t) + M(t), \quad (8)$$

with

$$Q(t) = \begin{bmatrix} BX(t) + \Gamma & & & \\ & \Sigma + \Gamma & & \\ & & D + \Gamma & \\ & & & A + \Gamma \end{bmatrix}, \quad (9)$$

$$M(t) = \begin{bmatrix} \Phi & 0 & 0 & A \\ BX(t) & \Phi & 0 & 0 \\ 0 & \Sigma & \Phi & 0 \\ 0 & 0 & D & \Phi \end{bmatrix}, \quad (10)$$

and let the complete state vector be defined as

$$z = [s^\top \ e^\top \ x^\top \ r^\top]^\top. \quad (11)$$

Thus, we can describe the state dynamics as

$$\dot{z} = (-Q(t) + M(t))z, \quad (12)$$

where $Q(t)$ is a positive diagonal matrix and $M(t)$ is a non-negative matrix $\forall t \geq 0$. Note that the time-dependence of $Q(t)$ and $M(t)$ comes from the fact that they are functions

of $X(t)$, where the indication of time dependence is left for emphasis of this fact.

Definition 1. A graph $\mathcal{G}(t) = (\mathbb{V}, \cup_{t \geq 0} \mathbb{E}(t), \mathbb{W})$ for $t \in \mathbb{R}_{\geq 0}$ is K -strongly connected if there exist some bound K such that $(\mathbb{V}, \cup_{j=t}^{t+K-1} \mathbb{E}(j), \mathbb{W})$ is strongly connected, for all $t \in \mathbb{R}_{\geq 0}$.

Assumption 3. Let the graph $\mathcal{G}(t) = (\mathbb{V}, \cup_{t \geq 0} \mathbb{E}(t), \mathbb{W})$, where $\mathbb{W} : \mathbb{E}(t) \rightarrow \mathbb{R}_{>0}$ is defined by $w_{ij}(t)$, be K -strongly connected.

Lemma 6. Given Assumptions 2-3, if $\exists t_0 \geq 0$ and $\exists i \in [n]$ such that $x_i(t_0) > 0$, the matrix $M(t_0)$ is irreducible.

Proof. Assume, by contradiction, that $M(t_0)$ is reducible. Then $M(t_0) - Q(t_0)$ will also be reducible, as $Q(t_0)$ is a diagonal matrix. Thus, the graph described by the adjacency matrix in (12) cannot be strongly connected. However, we can partition the graph into n strongly connected components where each component describes the evolution of the epidemic in each subpopulation. Consider the case where $n = 1$, which reduces to a non-networked SEIRS model. By Assumption 2, this system is strongly connected in the sense of the epidemic states given $x_1(t_0) > 0$ where

$$\begin{bmatrix} \dot{s}_1 \\ \dot{e}_1 \\ \dot{x}_1 \\ \dot{r}_1 \end{bmatrix} = \begin{bmatrix} -\beta_1 x_1(t_0) & 0 & 0 & \alpha_1 \\ \beta_1 x_1(t_0) & -\sigma_1 & 0 & 0 \\ 0 & \sigma_1 & -\delta_1 & 0 \\ 0 & 0 & \delta_1 & -\alpha_1 \end{bmatrix} \begin{bmatrix} s_1 \\ e_1 \\ x_1 \\ r_1 \end{bmatrix}.$$

For $n > 1$, the network connections are defined by the total flows between each sub-population. Thus, for the system to be reducible, there must exist a sub-population that is unreachable via travel from at least one other sub-population in the graph, which contradicts Assumption 3. Therefore, $M(t_0)$ must be irreducible. $\square$

Note that the irreducibility of $M(t)$ will play an important role in proving the existence of an endemic equilibrium, which will be proven in Section 4.

4 NETWORK FLOWS MODEL ANALYSIS

In this section, we analyze the fundamental behaviors of our model, which is divided into the analysis of the healthy and the endemic states of the system in (4).

4.1 Healthy State Analysis

We first consider the existence of a disease-free, or healthy state of the system and when the stability of the healthy state can be ensured. We then examine the effect of manipulating travel flows on the stability of the healthy state and prove that there is no permissible perturbation of flows that will change the local stability of the healthy state for any system.

Proposition 2. The healthy state, $(s^*, e^*, x^*, r^*) = (1, 0, 0, 0)$, is always an equilibrium of the system.

Proof. Substituting $(1, 0, 0, 0)$ into (3) yields

$$\dot{s}_i = \frac{1}{N_i} \sum_{j \neq i} (F_{ij} - F_{ji}), \quad \dot{e}_i = 0, \quad \dot{x}_i = 0, \quad \dot{r}_i = 0$$

for all $i \in [n]$, where $\sum_{j \neq i} (F_{ij} - F_{ji}) = 0$ by Assumption 3. Thus, $(s^*, e^*, x^*, r^*) = (1, 0, 0, 0)$ is always an equilibrium of the system given Assumptions 1 and 2. $\square$

We now consider the stability of the healthy state of the system and the conditions needed to reach this equilibrium.

First, since $s_i + e_i + x_i + r_i = 1$ for all $i \in [n]$, we can rewrite the dynamics of (6) with respect to the exposed, infected, and recovered states as

$$\begin{bmatrix} \dot{e} \\ \dot{x} \\ \dot{r} \end{bmatrix} = \underbrace{\begin{bmatrix} -\Sigma - \Gamma + \Phi & B(I - E - X - R) & 0 \\ \Sigma & -D - \Gamma + \Phi & 0 \\ 0 & D & -A - \Gamma + \Phi \end{bmatrix}}_{\bar{H}} \begin{bmatrix} e \\ x \\ r \end{bmatrix}, \quad (13)$$

where $E = \text{diag}(e_i)$, $X = \text{diag}(x_i)$, and $R = \text{diag}(r_i)$. The Jacobian matrix of (13) evaluated at an equilibrium (x^*, e^*, r^*) , denoted as $J_{(e^*, x^*, r^*)}$, is

$$\begin{bmatrix} -\Sigma - \Gamma + \Phi - BX^* & B(I - E^* - 2X^* - R^*) & -BX^* \\ \Sigma & -D - \Gamma + \Phi & 0 \\ 0 & D & -A - \Gamma + \Phi \end{bmatrix}. \quad (14)$$

Further, we define the following matrix, the top 2x2 diagonal block of J evaluated at $(0, 0, 0)$,

$$U = \begin{bmatrix} -\Sigma - \Gamma + \Phi & B \\ \Sigma & -D - \Gamma + \Phi \end{bmatrix}, \quad (15)$$

which yields the following conditions for the local stability and instability of the healthy state.

Proposition 3. *If $s(U) < 0$, then the healthy state of (6) is locally exponentially stable.*

Proof. Evaluating (14) at the healthy state equilibrium yields

$$J_{(1,0,0,0)} = \begin{bmatrix} -\Sigma - \Gamma + \Phi & B & 0 \\ \Sigma & -D - \Gamma + \Phi & 0 \\ 0 & D & -A - \Gamma + \Phi \end{bmatrix}. \quad (16)$$

Since (16) is block lower triangular, we can examine the eigenvalues of each block separately. By Assumption 1 we have that $\gamma_i = \sum_{j \neq i} \phi_{ij}$, thus by the Gershgorin Circle Theorem we have that $s(-A - \Gamma - \Phi) < 0$. The upper block is equal to U , which is negative definite by assumption. Thus, we have that $s(J_{(1,0,0,0)}) < 0$, which satisfies Lyapunov's indirect method for determining the local stability of the healthy state. $\square$

Corollary 1. *If $s(U) > 0$, then the healthy state of (6) is unstable.*

Proof. This result follows by the same method as the proof of Proposition 3, which leverages (14) evaluated at the healthy state, shown in (16), which is block lower triangular. Since the upper block is equal to U , if $s(U) > 0$ then $s(J_{(1,0,0,0)}) > 0$. Thus, by Lyapunov's indirect method, the healthy state is unstable. $\square$

We now discuss the local stability of the healthy state for (13) with respect to changes in the rate of flow between nodes.

Theorem 1. *If $J_{(1,0,0,0)}$, as defined in (16), has distinct eigenvalues then, given Assumptions 1 and 2, there exists no permissible perturbation to the population flows γ_i , for $i \in [n]$, that will change the local stability of the healthy state of (6).*

Proof. Let γ'_i be some perturbation to γ_i , $i \in [n]$ such that Assumption 1 is satisfied and $\gamma'_i = \gamma_i + \theta_i$. We can then define the perturbed system dynamics of (13) with the matrix

$$\bar{H}' = \begin{bmatrix} -\Sigma - \Gamma' + \Phi' & B(I - E - X - R) & 0 \\ \Sigma & -D - \Gamma' + \Phi' & 0 \\ 0 & D & -A - \Gamma' + \Phi' \end{bmatrix}, \quad (17)$$

where $\Gamma' = \Gamma + \Theta$, $\Theta = \text{diag}(\theta_i)$, and $\Phi' = N^{-1}W(\Gamma + \Theta)N$. We can then separate the perturbed system into

$$\bar{H}' = \bar{H} + \bar{E}, \quad (18)$$

where $\bar{H}$ is the original system dynamics defined in (13) and

$$\bar{E} = \begin{bmatrix} N^{-1}W\Theta N - \Theta & 0 & 0 \\ 0 & N^{-1}W\Theta N - \Theta & 0 \\ 0 & 0 & N^{-1}W\Theta N - \Theta \end{bmatrix}. \quad (19)$$

Taking the Jacobian of (18) yields

$$J'_{(e^*, x^*, r^*)} = J_{(e^*, x^*, r^*)} + \bar{E}.$$

We consider the healthy state equilibrium, which exists for all allowed parameters of our system by Proposition 2. Since the eigenvalues of $J_{(1,0,0,0)}$ are distinct, we have that $J_{(1,0,0,0)}$ is diagonalizable by some matrix $\bar{S}$, where $J_{(1,0,0,0)} = \bar{S}\Lambda\bar{S}^{-1}$ and $\Lambda = \text{diag}(\lambda_i)$. Thus, by Lemma 4 we have that if λ' is an eigenvalue of J' , then there exists some eigenvalue λ_i of J such that

$$|\lambda' - \lambda_i| \leq \kappa(\bar{S})_\infty \|\bar{E}\|_\infty. \quad (20)$$

However, by Assumption 1 we have that $\theta_i = \sum_{j \neq i} \frac{N_j}{N_i} w_{ij} \theta_j$. Thus, $\|\bar{E}\|_\infty = 0$ and for any eigenvalue λ' of J' there exists some eigenvalue λ_i , $i \in [n]$ such that $\lambda' = \lambda_i$. Therefore, no perturbation of the travel flows that maintains Assumption 1, and thus keeps the model well-defined, can change the local stability of the healthy state of (6). $\square$

4.2 Endemic State Analysis

We now present conditions for when an endemic state of the system, where $z^* \gg 0$, will exist. For clarity, we define the following matrices explicitly as a function of the full system state

$$Q^*(z^*) = \begin{bmatrix} BX^*(z^*) + \Gamma & 0 & 0 & 0 \\ 0 & \Sigma + \Gamma & 0 & 0 \\ 0 & 0 & D + \Gamma & 0 \\ 0 & 0 & 0 & A + \Gamma \end{bmatrix} \quad (21)$$

$$M^*(z^*) = \begin{bmatrix} \Phi & 0 & 0 & A \\ BX^*(z^*) & \Phi & 0 & 0 \\ 0 & \Sigma & \Phi & 0 \\ 0 & 0 & D & \Phi \end{bmatrix}, \quad (22)$$

where $X^*(z^*) = \text{diag}([0 \ 0 \ I_n \ 0] z^*)$.

Theorem 2. *If $\inf_{t \geq t_0} s(-Q(t) + M(t)) > 0$, then there exists an endemic equilibrium such that $z^* \gg 0$.*

Proof. Define the continuous map $f : (0, 1]^{4n} \rightarrow [0, b]^{4n}$, where $b \in \mathbb{R}_{\geq 1}$, given by

$$f(z) = Q^*(z)^{-1} M^*(z) z. \quad (23)$$

Since the domain of f is $(0, 1]^{4n}$, z as an argument of f satisfies $z \gg 0$. Further, by Lemma 6, since $Q^*(z)^{-1} M^*(z)$ is an irreducible, non-negative matrix for all $z \gg 0$, by Lemma 2, there exists a $v \gg 0$ such that

$$Q^*(z)^{-1} M^*(z) v = cv, \quad (24)$$

where $c = \rho(Q^*(z)^{-1} M^*(z))$. Since $\inf_{t \geq t_0} s(-Q(t) + M(t)) > 0$ we have, by Proposition 1, that $c > 1$. We can then find an $\varepsilon > 0$ such that for each $i \in [4n]$

$$\varepsilon v_i \leq \frac{c-1}{c} \quad (25)$$

and

$$\varepsilon v_i \leq z_i. \quad (26)$$

From (25), it follows that $1 \leq \frac{c}{1+\varepsilon cv_i}$ and thus $\varepsilon v_i \leq \frac{\varepsilon cv_i}{1+\varepsilon cv_i}$, yielding

$$\varepsilon v_i \leq \frac{(Q^*(z)^{-1} M^*(z) \varepsilon v)_i}{1 + (Q^*(z)^{-1} M^*(z) \varepsilon v)_i}. \quad (27)$$

We can show that the right hand side of (27) is less than or equal to $f(\varepsilon v)_i = (Q^*(\varepsilon v)^{-1} M^*(\varepsilon v) \varepsilon v)_i$ for each $i \in [4n]$ as follows. We can expand (23) using (11) as

$$f(z) = \begin{cases} (diag(Bx^z) + \Gamma)^{-1} (Ar^z + \Phi s^z) & (28a) \\ (\Sigma + \Gamma)^{-1} (diag(Bx^z) s^z + \Phi e^z) & (28b) \\ (D + \Gamma)^{-1} (\Sigma e^z + \Phi x^z) & (28c) \\ (A + \Gamma)^{-1} (Dx^z + \Phi r^z) & (28d) \end{cases}$$

where we can write $z = [s^z \top \ e^z \top \ x^z \top \ r^z \top]^\top$. Note that for (28b)-(28d), if we increase any value z_i the output also increases, i.e., (28b)-(28d) are monotonic. Thus, since $\varepsilon v_i \leq z_i$, for any $f(\varepsilon v)_i$ in (28b)-(28d), we have $\varepsilon v_i \leq f(\varepsilon v)_i$.

We now consider (28a), which is not monotonic, and show that $\varepsilon v_i \leq f(\varepsilon v)_i$, for the last case (28a). To show this is true, we write the vector v as

$$v = [v^s \top \ v^e \top \ v^x \top \ v^r \top]^\top$$

and note that by (26) we have

$$\beta_i \varepsilon v_i^x + \gamma_i \leq \beta_i x_i + \gamma_i, \quad (29)$$

for every $i \in [n]$. To simplify notation, let

$$a_i = (\alpha_i v_i^r + (\Phi v^s)_i). \quad (30)$$

We then have that

$$\beta_i \varepsilon v_i^x + \gamma_i \leq (\beta_i x_i + \gamma_i)(1 + (\beta_i x_i + \gamma_i)^{-1} \varepsilon a_i) \quad (31)$$

since $(\beta_i x_i + \gamma_i)^{-1} \varepsilon a_i \geq 0$, which then yields

$$\frac{(\beta_i x_i + \gamma_i)^{-1}}{1 + (\beta_i x_i + \gamma_i)^{-1} \varepsilon a_i} \leq (\beta_i \varepsilon v_i^x + \gamma_i)^{-1}. \quad (32)$$

Finally, multiplying both sides of (32) by εa_i we have

$$\frac{(\beta_i x_i + \gamma_i)^{-1} \varepsilon a_i}{1 + (\beta_i x_i + \gamma_i)^{-1} \varepsilon a_i} \leq (\beta_i \varepsilon v_i^x + \gamma_i)^{-1} \varepsilon a_i. \quad (33)$$

Therefore, since by (27) we have

$$\varepsilon v_i^s \leq \frac{(\beta_i x_i + \gamma_i)^{-1} \varepsilon (\alpha_i v_i^r + (\Phi v^s)_i)}{1 + (\beta_i x_i + \gamma_i)^{-1} \varepsilon (\alpha_i v_i^r + (\Phi v^s)_i)}$$

and by (28a)

$$f(\varepsilon v^s)_i = (\beta_i \varepsilon v_i^x + \gamma_i)^{-1} \varepsilon (\alpha_i v_i^r + (\Phi v^s)_i),$$

we have that $\varepsilon v_i^s \leq f(\varepsilon v^s)_i$, for all $i \in [n]$, by (33). Thus, we have shown that $\varepsilon v_i \leq f(\varepsilon v)_i$, for all $i \in [4n]$.

We now show that $f(z)$ is bounded by a $b \in \mathbb{R}$. Since $0 \ll z \leq 1$ we have by Assumption 2 and (28) that $f(z)$ must also be finite and we can choose $b = \sup_{i \in [4n]} f(z)_i < \infty$. Thus, f maps the convex compact set $\mathcal{C} = \{z \mid \varepsilon v \leq z \leq 1b\}$ to itself. By Brouwer's fixed-point theorem, f has a fixed point in $\mathcal{C}$, which must be strictly positive. Let z^* be this fixed point, then $f(z^*) = z^*$, i.e.

$$z^* = Q^*(z^*)^{-1} M^*(z^*) z^*. \quad (34)$$

Therefore, we have

$$\begin{aligned} Q^*(z^*) z^* &= M^*(z^*) z^* \\ 0 &= (M^*(z^*) - Q^*(z^*)) z^* \end{aligned}$$

and thus $z^* \gg 0$ is an equilibrium point of (4). Further, by Lemma 5 we know that $z^* \in (0, 1]^{4n}$. $\square$

Note that since our method of proving existence of an endemic equilibrium for (4) requires irreducibility we must include the susceptible state in our analysis, causing the sufficient condition needed to prove existence to be stronger than one that is time invariant. We now give conditions under which the endemic equilibrium is unique.

Theorem 3. *Let there exist an endemic equilibrium for the system in (4). If $\beta_i \geq \gamma_i$, for all $i \in [n]$, then the endemic equilibrium is unique.*

Proof. Let z and y be two nonzero equilibria of (4) which can be divided into their respective epidemic states as $z = [s^z \top \ e^z \top \ x^z \top \ r^z \top]^\top$ and $y = [s^y \top \ e^y \top \ x^y \top \ r^y \top]^\top$. We can express z in terms of y for $i \in [n]$ as

$$\begin{aligned} (s_i^y + \varepsilon_{s_i}) + (e_i^y + \varepsilon_{e_i}) + (x_i^y + \varepsilon_{x_i}) + (r_i^y + \varepsilon_{r_i}) \\ = s_i^z + e_i^z + x_i^z + r_i^z = 1, \end{aligned} \quad (35)$$

where

$$\varepsilon_{s_i} + \varepsilon_{e_i} + \varepsilon_{x_i} + \varepsilon_{r_i} = 0, \quad (36)$$

for each $i \in [n]$. Since both $z \gg 0$ and $y \gg 0$ are fixed points of (4), we can sum (4b), (4c), and (4d) to obtain

$$\begin{aligned}
\dot{e}_i^z + \dot{x}_i^z + \dot{r}_i^z &= \beta_i x_i^z s_i^z - \gamma_i (e_i^z + x_i^z + r_i^z) \\
&\quad - \alpha_i r_i^z + \sum_{j \neq i} \phi_{ij} (e_j^z + x_j^z + r_j^z) \\
&= \beta_i (x_i^y + \varepsilon_{x_i}) (s_i^y + \varepsilon_{s_i}) - \alpha_i (r_i^y + \varepsilon_{r_i}) \\
&\quad - \gamma_i (e_i^y + x_i^y + r_i^y + \varepsilon_{e_i} + \varepsilon_{x_i} + \varepsilon_{r_i}) \\
&\quad + \sum_{j \neq i} \phi_{ij} (e_j^y + x_j^y + r_j^y + \varepsilon_{e_j} + \varepsilon_{x_j} + \varepsilon_{r_j}) \\
&= \beta_i x_i^y s_i^y - \gamma_i (e_i^y + x_i^y + r_i^y) \\
&\quad - \alpha_i r_i^y + \sum_{j \neq i} \phi_{ij} (e_j^y + x_j^y + r_j^y) \\
&= \dot{e}_i^y + \dot{x}_i^y + \dot{r}_i^y = 0,
\end{aligned}$$

where $\phi_{ij} = \frac{N_j}{N_i} w_{ij} \gamma_j$. Thus, by taking the difference of $(\dot{e}_i^z + \dot{x}_i^z + \dot{r}_i^z) - (\dot{e}_i^y + \dot{x}_i^y + \dot{r}_i^y)$ we get

$$\begin{aligned}
&\beta_i (x_i^y \varepsilon_{s_i} + s_i^y \varepsilon_{x_i} + \varepsilon_{x_i} \varepsilon_{s_i}) - \gamma_i (\varepsilon_{e_i} + \varepsilon_{x_i} + \varepsilon_{r_i}) \\
&\quad - \alpha_i \varepsilon_{r_i} + \sum_{j \neq i} \phi_{ij} (\varepsilon_{e_j} + \varepsilon_{x_j} + \varepsilon_{r_j}) = 0.
\end{aligned} \tag{37}$$

We can express (37) for all $i \in [n]$ using matrix notation as follows. Define $\varepsilon_q = [\varepsilon_{q_1}, \dots, \varepsilon_{q_n}]^\top$, where $q \in \{s, e, x, r\}$. Then, using (36) and (37) we have

$$B(X^y \varepsilon_s + S^y \varepsilon_x + E_x \varepsilon_s) + \Gamma \varepsilon_s - A \varepsilon_r - \Phi \varepsilon_s = \mathbf{0}, \tag{38}$$

where $X^y = \text{diag}(x_i^y)$, $S^y = \text{diag}(s_i^y)$, and $E_x = \text{diag}(\varepsilon_{x_i})$. Further, we can put (38) into a block matrix form as

$$\underbrace{\begin{bmatrix} BX^y + E_x + \Gamma - \Phi & 0 & 0 \\ 0 & BS^y & 0 \\ 0 & 0 & -A \end{bmatrix}}_K \begin{bmatrix} \varepsilon_s \\ \varepsilon_x \\ \varepsilon_r \end{bmatrix} = \mathbf{0}. \tag{39}$$

Thus, the solution to the vector of perturbations to s^y, x^y, r^y is the nullspace of K . Note that if K is full rank, then the only solution for the perturbations must be $\varepsilon_s = \varepsilon_x = \varepsilon_r = \mathbf{0}$. Considering the rank of K , it is clear that the second and third block diagonals are full rank since $y \gg 0$ and by Assumption 2, respectively. Further, by Lemma 3, we have that if $BX^y + E_x + \Gamma - \Phi$ is diagonally dominant then it is also full rank, or in other words, if

$$|\beta_i x_i^y + \varepsilon_{x_i} + \gamma_i| > \sum_{j \neq i} |\phi_{ij}|, \quad \forall i \in [n], \tag{40}$$

then $\det(BX^y + E_x + \Gamma - \Phi) \neq 0$.

We will now show that (40) holds for any choice of ε_{x_i} if $\beta_i \geq 1$ for all $i \in [n]$. Note that when $\varepsilon_{x_i} \geq 0$ we have that (40) is satisfied, since $\gamma_i = \sum_{j \neq i} \phi_{ij}$ for all $i \in [n]$ by Assumption 1, and $\beta_i x_i^y > 0$ by Assumption 2. We now consider when $\varepsilon_{x_i} < 0$. Since $x_i^y + \varepsilon_{x_i} = x_i^z > 0$, we have that $x_i^y > -\varepsilon_{x_i}$. Therefore, if $\varepsilon_{x_i} < 0$, then we have that $x_i^y > |\varepsilon_{x_i}|$. Thus, if $\beta_i \geq 1$ for all $i \in [n]$, then (40) must hold for any feasible choice of ε_{x_i} . Therefore, by (39) we have $\varepsilon_s = \varepsilon_x = \varepsilon_r = \mathbf{0}$, which implies, by (36), that $\varepsilon_e = \mathbf{0}$. Thus, by (35), $z = y$.

Now consider a system with an endemic equilibrium where $\exists i \in [n]$ such that $\beta_i < 1$. If $\beta_i \geq \gamma_i$ for all $i \in [n]$, there exists some constant scalar $\eta \in \mathbb{R}$ such that by (6),

$$\begin{aligned}
\eta (A r^* - (BX^* + \Gamma) s^* + N^{-1} W \Gamma N s^*) &= \mathbf{0} \\
\eta (BX^* s^* - (\Sigma + \Gamma) e^* + N^{-1} W \Gamma N e^*) &= \mathbf{0},
\end{aligned}$$

where $\eta \beta_i \geq 1$ and $\eta \gamma_i \leq 1$ for all $i \in [n]$. Thus, any endemic system with parameters where $\exists i \in [n]$ such that $\beta_i < 1$ can be mapped to one with the same endemic equilibrium where $\beta_i \geq 1$ for all $i \in [n]$. $\square$

Remark 1. It should be noted that the condition $\beta_i \geq \gamma_i, \forall i \in [n]$, for uniqueness of the endemic equilibrium is only sufficient and simulations show that an endemic equilibrium can still exist even when $\exists i \in [n]$ such that $\beta_i < \gamma_i$.

The analysis throughout this section establishes several key points. First, we find that short of a complete lockdown of infected nodes before spreading begins, manipulating travel flows alone cannot change the long-term stability of the healthy state of this system. Further, we have established conditions under which the healthy state is exponentially stable, the healthy state is unstable, an endemic equilibrium exists, and the endemic equilibrium is unique.

5 APPLICATIONS OF TRAVEL FLOWS MODEL

In this section, we shift our focus to constructing novel tools that will enable us to leverage real-world data in Section 6. The proposed tools are 1) learning the infection spread parameters given infection states and travel flows data and 2) using travel flows data to predict the arrival time of a disease to every node in the network given a disease origin. In this section, we illustrate these tools using simulated data. We share the simulation code used for this section on GitHub at [52].

5.1 Parameter Identification

In this section we present a method for estimating the infection parameters $\beta_i, \sigma_i, \delta_i, \alpha_i$ for each node i , respectively, given measurement data. Since infection data is generally collected at discrete time intervals, we first discretize the system in (4) using Euler's method which yields

$$s_i^{k+1} = s_i^k + h \left(\alpha_i r_i^k - (\beta_i x_i^k + \gamma_i^k) s_i^k + \sum_{j \neq i} \frac{N_j}{N_i} w_{ij}^k \gamma_j^k s_j^k \right) \tag{41a}$$

$$e_i^{k+1} = e_i^k + h \left(\beta_i x_i^k s_i^k - (\sigma_i + \gamma_i^k) e_i^k + \sum_{j \neq i} \frac{N_j}{N_i} w_{ij}^k \gamma_j^k e_j^k \right) \tag{41b}$$

$$x_i^{k+1} = x_i^k + h \left(\sigma_i e_i^k - (\delta_i + \gamma_i^k) x_i^k + \sum_{j \neq i} \frac{N_j}{N_i} w_{ij}^k \gamma_j^k x_j^k \right) \tag{41c}$$

$$r_i^{k+1} = r_i^k + h \left(\delta_i x_i^k - (\alpha_i + \gamma_i^k) r_i^k + \sum_{j \neq i} \frac{N_j}{N_i} w_{ij}^k \gamma_j^k r_j^k \right), \tag{41d}$$

where $h > 0$ is a sampling parameter and $k \in \mathbb{Z}_{\geq 0}$ is a given time index. For this application, we assume a sampling parameter h that is small enough such that the system remains well-defined given the relative values of the model parameters. For a more detailed discussion and analysis of exactly when this condition holds, see our previous work in [47]. Given T samples of infection state data, we can estimate the parameters for node i as follows. Let

$$\Delta q_i = \begin{bmatrix} q_i^1 - q_i^0 + h(\gamma_i^0 q_i^0 - \sum_{j \neq i} \frac{N_j}{N_i} w_{ij}^0 \gamma_j^0 q_j^0) \\ \vdots \\ q_i^T - q_i^{T-1} + h(\gamma_i^{T-1} q_i^{T-1} - \sum_{j \neq i} \frac{N_j}{N_i} w_{ij}^{T-1} \gamma_j^{T-1} q_j^{T-1}) \end{bmatrix} \quad (42)$$

where $q_i \in \{s_i, e_i, x_i, r_i\}$, and define

$$\Psi_i = h \begin{bmatrix} -s_i^0 x_i^0 & 0 & 0 & r_i^0 \\ \vdots & \vdots & \vdots & \vdots \\ -s_i^{T-1} x_i^{T-1} & 0 & 0 & r_i^{T-1} \\ s_i^0 x_i^0 & -e_i^0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ s_i^{T-1} x_i^{T-1} & -e_i^{T-1} & 0 & 0 \\ 0 & e_i^0 & -x_i^0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & e_i^{T-1} & -x_i^{T-1} & 0 \\ 0 & 0 & x_i^0 & -r_i^0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & x_i^{T-1} & -r_i^{T-1} \end{bmatrix}, \quad (43)$$

then, we rewrite the system in (41) for a node i given T data samples of data in terms of the spread parameters as

$$\begin{bmatrix} \Delta s_i \\ \Delta e_i \\ \Delta x_i \\ \Delta r_i \end{bmatrix} = \Psi_i \begin{bmatrix} \beta_i \\ \sigma_i \\ \delta_i \\ \alpha_i \end{bmatrix}. \quad (44)$$

Thus, we can solve for the parameters $\beta_i, \sigma_i, \delta_i, \alpha_i$ as

$$\begin{bmatrix} \beta_i \\ \sigma_i \\ \delta_i \\ \alpha_i \end{bmatrix} = \Psi_i^\dagger \begin{bmatrix} \Delta s_i \\ \Delta e_i \\ \Delta x_i \\ \Delta r_i \end{bmatrix}, \quad (45)$$

where $\Psi_i^\dagger$ is the pseudo-inverse of Ψ_i . Thus, we have that a solution to the spread parameters using (45) is unique if and only if (43) is full column rank.

Using this method of parameter estimation, we can utilize real infection data along with real travel flows data to determine the spreading rates for a given disease outbreak, assuming that the model defined in Section 3 approximates the behavior of the outbreak. We show an example of parameter identification in Figure 1, where data from a 5-node simulated system with initial conditions $s^0 = [0.549, 0.715, 0.603, 0.545, 0.424]^\top$, $e^0 = \mathbf{1} - s^0$, and $x^0 = r^0 = \mathbf{0}$ is perturbed with Gaussian noise with zero mean and standard deviation 0.01 and then used to learn the infection spread parameters as outlined in this section, where the travel weight matrix used is given in Table 1 and the spread parameters and of the learned parameters are given in Table 2.

w_{ij}	1	2	3	4	5
1	0	0.212	0.275	0.25	0.212
2	0.249	0	0.26	0.299	0.338
3	0.246	0.198	0	0.204	0.178
4	0.285	0.29	0.259	0	0.272
5	0.22	0.299	0.206	0.247	0

TABLE 1: Travel connection matrix used in the simulated system shown in Figure 1.

Node	1	2	3	4	5	
β_i	0.065	0.044	0.089	0.096	0.038	0.03
σ_i	0.079	0.053	0.057	0.093	0.007	0.01
δ_i	0.001	0.001	0.008	0.008	0.009	0.002
α_i	0.01	0.008	0.005	0.008	0.001	0.012
γ_i	0.002	0.002	0.002	0.002	0.005	-

TABLE 2: Parameters used in the simulated system shown in Figure 1 along with the of learned parameters.

5.2 Effective Distance and Travel Flows

Given our notion of population flow and the definition of w_{ij} being the proportion of traveling members of the population at node j flowing to node i , we can use this proportional flow information to compute the effective distance between any two nodes based on the most probable path an individual would take between nodes. Further, we can then use the effective distance between nodes to make predictions of the arrival time of an outbreak to a given node based on the distance of that node from other infected nodes. This notion of effective distance based on travel and its use in predicting disease arrival time is introduced in [32], [33], where the effective distance between countries is computed using flight traffic data. The work in [34] extends the notion of effective distance to incorporate multiple outbreak locations for the prediction of disease arrival times. In [35], travel flows and effective distance metrics are used to model higher-order effects resulting from population travel movements. Additionally,

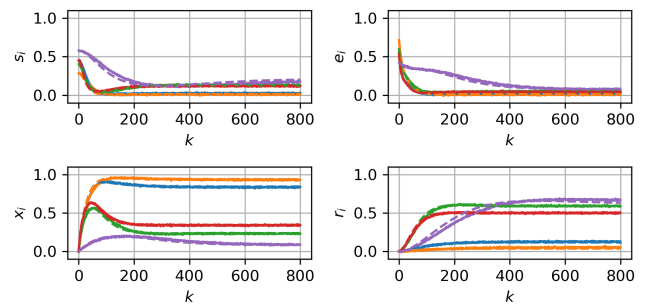


FIGURE 1: Simulated 5-node system from (41) with parameters and travel matrix defined in Tables 2 and 1 where $n = 5$, $h = 1$, and added zero-mean Gaussian noise with standard deviation 0.01 and spread parameters chosen to induce an endemic state. Each solid line represents the perturbed simulated data for each node. Parameters from the noisy system are identified and then re-simulated from the same initial condition, where the predicted state trajectories for each node are shown by dashed lines of the same color.

in [36] they use effective distance to create an infection delay model that estimates infection delays under multiple possible outbreak scenarios. In our work, we apply this method of using effective distance via travel flows to predict disease arrival time using to inter-county travel data in Minnesota during a period of the COVID-19 pandemic, which will be shown in Section 6.

We compute the effective distance between nodes as follows. We interpret the mobility matrix defined by the w_{ij} as the transition matrix of a discrete time homogeneous stochastic jump process, where at each time step a randomly moving particle moves from node j to node i with probability w_{ij} . Let n_k be the location of this particle at time k . Now, consider the path of a randomly moving particle after L steps, forming the path $\Xi = \{n_0, n_1, \dots, n_L\}$. If we fix any two endpoints in the graph where Ξ_{ij} is any path with $n_0 = j$ and $n_L = i$, we can associate any path taken between these points with a probability

$$F(\Xi_{ij}) = w_{in_{L-1}} \times \dots \times w_{n_1j} = \prod_{k=1}^L w_{n_k n_{k-1}}. \quad (46)$$

Since the proposed notion of distance is additive, we establish a connection between the probability of a given path and the effective distance of a given path by employing the multiplicative property of logarithms

$$D_F(\Xi_{ij}) = -\log F(\Xi_{ij}) = -\sum_{k=1}^L \log w_{n_k n_{k-1}}, \quad (47)$$

where $D_F(\Xi_{ij})$ is the log sum of the probabilities of each step taken on the given path. Thus, since $w_{ij} \in [0, 1]$ a low probability path will equate to a high effective distance, with $D_F(\Xi) \rightarrow \infty$ as $F(\Xi) \rightarrow 0$, and high probability paths result in lower effective distance. We construct a new graph with edge weights between nodes being the one-step negative log probabilities of traveling between nodes as

$$d_{ij} = \begin{cases} 0 & i = j \\ -\log w_{ij} & w_{ij} > 0 \\ \infty & w_{ij} = 0. \end{cases} \quad (48)$$

Using this graph with edge weights between nodes i and j defined by (48), we compute the effective distance as

$$D_{ij} = \min_{\Xi_{ij}} D_F(\Xi_{ij}) \geq 0, \quad (49)$$

where D_{ij} is the sum of the path lengths between nodes j and i that yield the smallest distance, which can be found for any node $i \in [n]$ by computing the minimum spanning tree for the distance graph from node j , since $D_F(\Xi_{ij}) = \sum_{k=1}^L d_{n_k n_{k-1}}$ describes the cost of traveling any path between i and j in the graph. It is important to note that since it is possible $d_{ij} \neq d_{ji}$, we also have that $D_{ij} \neq D_{ji}$. Further, the shortest path to and from any two endpoints may not necessarily contain the same intermediate points.

We can extend the notion of the most probable path and effective distance further by considering the most probable path from a subgroup of nodes to any given node. Since

we are interested in the effective distance for an infection to travel to uninfected nodes with respect to an infected group of nodes, we can construct a node subgroup as follows

$$\mathcal{X} = \{i \in [n]; x_i > p\}, \quad (50)$$

where x_i is the infection level at node i and $p \in (0, 1]$ is some threshold at which we consider the node to be infected. We can then compute the probability that an individual from node group $\mathcal{X}$ travels to node i as

$$\tilde{w}_{i\mathcal{X}} = \frac{1}{\sum_{l \in \mathcal{X}} N_l} \sum_{j \in \mathcal{X}} N_j w_{ij}. \quad (51)$$

This method incorporates the relative population size between nodes as well as the total likelihood that an individual from the infected node group $\mathcal{X}$ would travel to a given node i . Another method for computing the effective distance of a single node to a group of outbreak nodes is given in [34], which instead computes the effective distance to the infected group $\mathcal{X}$ independently of population size

$$\tilde{w}_{i\mathcal{X}} = \frac{1}{n} \sum_{j \in \mathcal{X}} \frac{1}{e^{d_{ij}}}, \quad (52)$$

which is analogous to the computation of the effective resistance in parallel circuits.

Thus, we can construct another effective distance graph, with respect to a group of nodes $\mathcal{X}$, using edge weights

$$\tilde{d}_{ij}^{\mathcal{X}} = \begin{cases} 0 & i \in \mathcal{X} \text{ or } i = j \\ -\log \tilde{w}_{i\mathcal{X}} & i \notin \mathcal{X}, j \in \mathcal{X}, \tilde{w}_{i\mathcal{X}} > 0 \\ -\log w_{ij} & i, j \notin \mathcal{X}, w_{ij} > 0 \\ \infty & \text{otherwise,} \end{cases} \quad (53)$$

making the effective distance in the case from any node j to every other node $i \in [n]$, with respect to the node group $\mathcal{X}$,

$$\tilde{D}_{ij}^{\mathcal{X}} = \min_{\Xi_{ij}} (\tilde{d}_{jn_{L-1}}^{\mathcal{X}} + \dots + \tilde{d}_{n_1 i}^{\mathcal{X}}). \quad (54)$$

When a disease outbreak occurs, one important question for policymakers is how long it will take for the disease to infect new nodes given past infection data. Given travel data and knowledge of which nodes are currently infected, we can use (54) to predict the arrival time of the disease to uninfected nodes as follows. Let T_k be the time of the k th arrival of the disease to an uninfected node, and suppose $k > \tau$ disease arrivals have occurred, where τ is the size of your training window:

- 1) At time T_k , select training data that includes the disease arrivals from time $T_{k-\tau}$ to T_k , where $\tau = |\{T_{k-\tau}, T_{k-\tau+1}, \dots, T_k\}|$
- 2) Compute the effective distance of each node in the training data to the infected group $\mathcal{X}$ at time $T_{k-\tau}$
- 3) Given the arrival times of the training data and the computed effective distances to the infected group, fit a line to these training data
- 4) Using the fitted line, predict the arrival time of the remaining uninfected nodes given their effective distance to the current infected group $\mathcal{X}$ at time T_k .

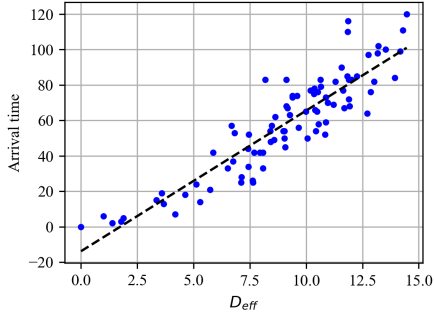
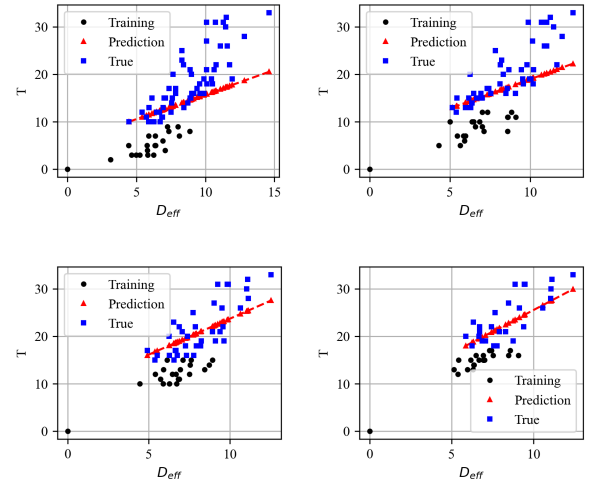


FIGURE 2: The arrival time of the disease to each node versus the effective distance of each node in our simulated system from the disease origin, computed by (49). The dashed line shows the linear fit of all data points.

To ensure that this method always predicts sensible arrival times, as the fitted line may predict arrival times that have already passed, a constant shift is added to the predictions such that the next predicted arrival time is greater than zero. This method differs from [32], [33] most significantly in its ability to make updated predictions on the arrival time of an outbreak, which depends on the current flow rate. Further, it facilitates a notion of the distance of an outbreak from any given node based on the likelihood of transmission from multiple source nodes as the epidemic progresses similarly to [34], described by (53), rather than relying on the distance from the original source. Note that (53) may use either the population-dependent model described in (51) or the population-independent model described in (52) as proposed by [34]. Additionally, our prediction algorithm may use any method of computing the effective distance to multiple infected nodes, so long as it accounts for the proportional travel flow between nodes and the set of currently infected nodes. Further, note that (53) cannot use the methods from [35] and [36] as they do not handle the case of multiple outbreak locations. These factors contribute to a more accurate prediction of arrival time in a more realistic evolving scenario.

We illustrate this algorithm by simulating the spread of a disease over a network using our model as defined in Section 3. We construct our network to emulate counties in Minnesota with travel between counties given by inter-county travel data collected during part of the COVID-19 pandemic. A full description of this data is given in Section 6. The system is then simulated using infection parameters satisfying Assumption 2 and sampling parameter $h = 1$ (to emulate a daily resolution of data collection), where a node is selected as the origin of the disease and predictions on the next nodes to be infected are made. Using (49), we compute the effective distance of each node from the origin of infection and plot the effective distance versus the arrival time of the disease to each node, shown in Figure 2. Note the linear relationship between effective distance and arrival time.

In Figure 3 we show an example of using the shifting window algorithm to predict the remaining infection arrival times with an arrival window of $\tau = 20$ disease



(C) $T = T_{70}$

(D) $T = T_{80}$

FIGURE 3: Example of shifting prediction window of arrival times in the simulated system at four different arrival times during the spreading process ($T = T_{50}, T_{80}$) where training window $\tau = 20$, with the data used to make predictions (black circles), true arrival times for each node versus effective distance (blue squares), and shifted best-fit predictions (red triangles) shown at several arrival intervals, where the effective distances are computed at each disease arrival using (54).

arrival events. To evaluate the prediction error of our shifting window method, we compute the root-mean-square error (RMSE) of the linear fit for all arrival times versus effective distance from the origin at the start of the epidemic (with the linear fit shown in Figure 2) and compare it with the average RMSE of the shifting window method that predicts the next 10 arrival times whenever a new arrival occurs. This comparison yields an RMSE of 4.59 timesteps (analogous to days since $h = 1$) for the linear fit of all arrival times versus effective distance from the disease origin (which can only be evaluated after all data has been collected) versus an average RMSE of 1.91 timesteps and 1.58 timesteps for the shifting window method using (51) and (52), respectively, that predicts the next 10 arrival times whenever a new arrival occurs, which is at least a 58% reduction in prediction error in addition to being usable in real-time to predict the next immediate disease arrivals versus analysis after the fact. Note that in the case of the simulated data, using (52) to compute the distance of uninfected nodes to the infected groups yields a marginally lower for predicting future arrival times compared with (51), as shown in Table 4; however, it is uncertain that (52) will outperform (51) in all cases. Further, (51) may provide a more explainable justification for estimating the effective distance between the infected group and uninfected nodes based on relative population sizes and the number of individuals traveling from the infected group to uninfected nodes.

6 CASE STUDY: MINNESOTA AND COVID-19

We apply our model developed in Section 3 and methods described in Section 5 to data collected from the beginning of the 2020 COVID-19 pandemic in Minnesota. Daily infection data at the county level is taken from usafacts.org which contains the number of confirmed COVID-19 cases over the period of March 2020 to December 2020 [53]. To infer the infection states s_i, e_i, x_i, r_i for each county i from the confirmed cases data we apply a constant shift that assumes confirmed individuals became exposed one week before testing positive, at which point they are considered infected. Then, we assume that infected individuals remain infected for one week after testing positive, after which they are considered recovered. Finally, to account for the rate at which individuals can lose immunity, we transfer recovered individuals back into being susceptible after six weeks of being recovered. For travel data, we worked with the Minnesota Department of Transportation to obtain data collected by StreetLight [54] that estimates the number of trips between counties via anonymous geolocalization using smartphones. This provides an estimate of the total number of trips made by individuals between counties over a specified period of time. We choose a weekly time scale in an effort to average out periodic behaviors such as commuting and use this average to estimate the daily flow of individuals between counties.

6.1 Parameter Identification

Using the inferred infection states for each county in MN along with the travel flows computed from the collected traffic data, we can use (45) to solve for the best-fit infection parameters using a convex solver and constraining solutions to be strictly non-negative using our model defined in Section 3. We then re-simulate the model using the same initial conditions and learned parameters, with the results of the simulation compared with the inferred infection states shown in Figure 4, where we show a subset of real infection levels at five counties (solid lines) and the predicted infection levels from the learned parameters (dashed lines). Note that although the predicted infection peak of the model with learned parameters is much higher than the actual data, we see that the peak infection time is more closely approximated by the travel flows model, as shown in Table 3.

Regarding the error in magnitude prediction, there are several likely contributing factors. First, in our estimation algorithm, we assume spread parameters ($\beta_i, \sigma_i, \delta_i, \alpha_i$) that are not time-varying, which is most likely not the case in reality since lock-downs, social distancing policies, medical interventions, and virus mutations can all change the spread parameters over time, which violates this assumption. Additionally, we are assuming that our rather naive method for state estimation (by choosing a priori the recovering rate and the rate at which exposed individuals become infected) yields the correct states for all counties and that all counties are reporting infection cases in the same manner, which may not be true in all cases. Thus, these combinations of structural assumptions and naivety

	Infection	Peak Time
SEIRS (Travel Flows)	0.023	4.15 Weeks
SEIRS ([46])	0.007	21.21 Weeks

TABLE 3: The infection $\bar{x}$, defined as $\sqrt{\frac{1}{nT} \sum_{i \in [n], k \in [T]} (x_i^k - \hat{x}_i^k)^2}$, where n is the number of nodes, T is the number of samples, and $\hat{x}_i^k$ is the predicted infection level at node i for time step k , and the peak time $\bar{t}$, defined as $\sqrt{\frac{1}{n} \sum_{i \in [n]} (\arg \max_k (x_i^k) - \arg \max_k (\hat{x}_i^k))^2}$, for the parameter fitting predictions of the SEIRS travel flows model defined in Section 3 (shown in Figure 4) and modified SEIRS model from [46] (shown in Figure 5), respectively, on the COVID-19 infection data.

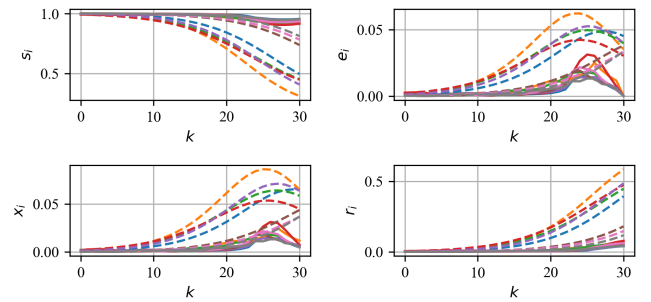


FIGURE 4: Simulating (4) for select MN counties using traffic data to estimate inter-county travel flows (dashed lines) with best-fit parameters learned from the estimated epidemic states using COVID-19 infection data (solid lines).

in state estimation are both likely large contributors to the magnitude error in state prediction.

For comparison, we apply this parameter fit with another similar model, modified from our previous work in [46] to allow individuals to become reinfected. The major difference between our model in Section 3 and this modified model from [46] is that travel data is used to infer direct interaction between sub-populations rather than explicitly modeling the flow of individuals between subpopulations. We then take the best-fit parameters for this model given the infection and travel data and simulate the model with the learned parameters, shown in Figure 5, where we show a subset of real infection levels at the same five counties as Figure 4 (solid lines) and the predicted infection levels from the learned parameters (dashed lines). Note that in this case, although the error is smaller in magnitude than the prediction from the parameter fitting of the travel flows model, the direct interaction model from [46] fails to reproduce the peaking behavior shown clearly in the inferred infection states from the COVID-19 data and in Table 3.

6.2 Predicting Disease Arrival Time

We use the methods described in Section 5.2 and the collected travel flows data to make predictions on the arrival time of COVID-19 to counties in MN. In Figure 6 we compare the arrival time of the first reported case with both the effective distance and geographic distance of each county to the disease origin (i.e., the county with

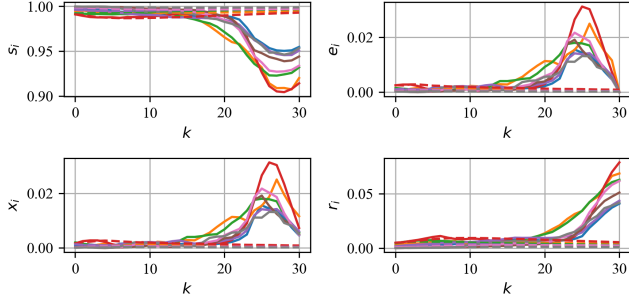


FIGURE 5: Simulating the modified direct connection model from [46] for select MN counties using traffic data to estimate inter-county travel flows (dashed lines) with best-fit parameters learned from the estimated epidemic states using COVID-19 infection data (solid lines).

	Arrival Time Prediction (days)		
	Shifting Window		[32], [33]
$\tilde{w}_{i,\mathcal{X}}$	(51)	(52) [34]	w_{ij}
Simulated Data	1.91	1.58	4.59
MN Data	4.02	3.68	18.86

TABLE 4: A comparison of the arrival time prediction for the shifting window prediction method with multiple outbreak locations as described in Section 5.2 and the single outbreak case of [32], [33].

the first reported case in MN), with the best-fit line shown for each. We compute the R-value for each best-fit line to be 0.53 and 0.48 for the effective distance and geographic distance, respectively, suggesting that the effective distance is a better predictor of arrival time, albeit by a small margin, than geographic distance when considering all arrival times with respect to the disease origin. Finally, we demonstrate the shifting window prediction method on the MN COVID-19 data in Figure 7. We find an RMSE of close to 19 days for the linear fit [32], [33] of all arrival times versus effective distance from the disease origin versus an average RMSE of 4.02 days and 3.68 days for the shifting window method using (51) and (52), respectively, that predicts the next 10 arrival times whenever a new arrival occurs, yielding at least a 79% reduction in prediction error in addition to enabling real-time arrival predictions. Again, note that in the case of the MN COVID-19 data, using (52) to compute the distance of uninfected nodes to the infected groups yields a marginally lower for predicting future arrival times compared with (51), as shown in Table 4.

7 CONCLUSION

In this work, we have developed a model for simulating the spread of an SEIRS epidemic process using networked population flows as the propagation mechanism. We provide sufficient conditions under which the healthy state of the system will be locally stable or unstable and show via model analysis that there exists no valid perturbation to the population flows that will change the local stability of any healthy state. These results suggest that manipulating

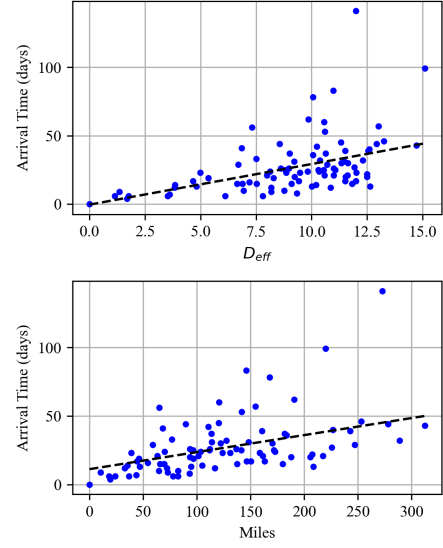
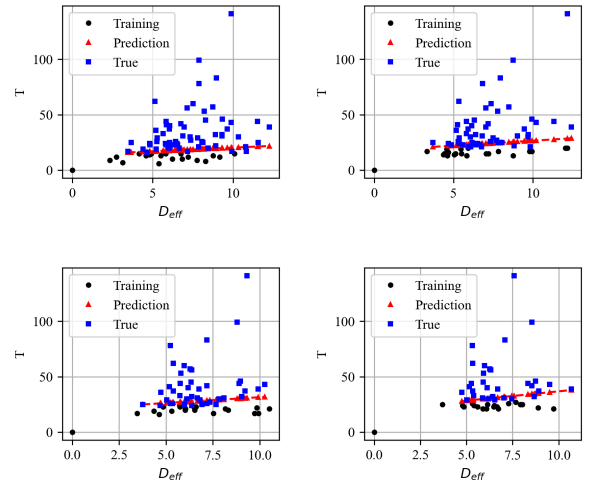


FIGURE 6: The arrival time of the disease to each node versus the effective distance (top) of each county in MN from the county with the first detected COVID-19 cases during the 2020 pandemic, computed by (49), compared with the geographic distance of each county (bottom) from the disease origin. The dashed line shows the linear fit of all data points for each case respectively.



(C) $T = T_{65}$

(D) $T = T_{75}$

FIGURE 7: Examples of the shifting prediction window of arrival times of COVID-19 for counties in MN at four different arrival times during the pandemic ($T = T_{45}, T_{75}$) where training window $\tau = 20$, with the data used to make predictions (black circles), true arrival times for each county versus effective distance (blue squares), and shifted best-fit predictions (red triangles) shown at several arrival intervals, where the effective distances is computed at each disease arrival using (54).

population flow alone is insufficient to achieve the long-term eradication of the disease in the defined system where a loss of immunity is present. Conversely, the result also suggests that perturbing the population flows will not prevent a system from achieving disease eradication (i.e., causing the healthy state to become unstable if it is already locally stable). Further, we provide sufficient conditions under which the system will enter a unique endemic state. Although the analytical results show that travel control is insufficient for affecting the long-term behavior of the epidemic, we show that travel flows are a good predictor of disease arrival time, which can provide valuable model-agnostic estimates of the arrival times for a given disease that are useful in allocating a fixed amount of medical resources. To this end, we then develop tools for applying our proposed model to data, such as spreading parameter identification and disease arrival time prediction and illustrate these tools using a case study of both travel and infection data from the counties in MN during part of the COVID-19 pandemic.

There are several directions for continued research on the work presented in this paper. Future directions for analysis included a study on the global stability of the healthy state, as only local stability is explored in this work, as well as a study of the stability of the endemic equilibrium for this system. Additionally, in this work, we have assumed that no infections occur during the transit of individuals between populations, and adding this element to our travel flows model may provide additional meaningful insights into the role of travel in spreading infectious diseases. Further, since the use of effective distance to predict disease arrival time is independent of the epidemic model being used, a more thorough analysis can be performed comparing the usefulness of effective distance in predicting arrival time for other epidemic models with travel flows (such as networked SIR, SIS, SAIR, etc.).

REFERENCES

- [1] S. Ruan, W. Wang, and S. A. Levin, "The effect of global travel on the spread of SARS," *Mathematical Biosciences & Engineering*, vol. 3, no. 1, p. 205, 2006.
- [2] A. J. Tatem, D. J. Rogers, and S. I. Hay, "Global transport networks and infectious disease spread," *Advances in Parasitology*, vol. 62, pp. 293–343, 2006.
- [3] J. Lowe, P. Gauger, K. Harmon, J. Zhang, J. Connor, P. Yeske, T. Loula, I. Levis, L. Dufresne, and R. Main, "Role of transportation in spread of porcine epidemic diarrhea virus infection, united states," *Emerging Infectious Diseases*, vol. 20, no. 5, p. 872, 2014.
- [4] A. Findlater and I. I. Bogoch, "Human mobility and the global spread of infectious diseases: a focus on air travel," *Trends in Parasitology*, vol. 34, no. 9, pp. 772–783, 2018.
- [5] T. D. Hollingsworth, N. M. Ferguson, and R. M. Anderson, "Will travel restrictions control the international spread of pandemic influenza?" *Nature Medicine*, vol. 12, no. 5, pp. 497–499, 2006.
- [6] M. Tang, Z. Liu, and B. Li, "Epidemic spreading by objective traveling," *EPL (Europhysics Letters)*, vol. 87, no. 1, p. 18005, 2009.
- [7] S. Maital and E. Barzani, "The global economic impact of COVID-19: A summary of research," *Samuel Neaman Institute for National Policy Research*, vol. 2020, pp. 1–12, 2020.
- [8] A. D. Kaye, C. N. Okeagu, A. D. Pham, R. A. Silva, J. J. Hurley, B. L. Arron, N. Sarfraz, H. N. Lee, G. E. Ghali, J. W. Gamble et al., "Economic impact of COVID-19 pandemic on healthcare facilities and systems: International perspectives," *Best Practice & Research Clinical Anaesthesiology*, vol. 35, no. 3, pp. 293–306, 2021.
- [9] T. Ahmad, M. Baig, and J. Hui, "Coronavirus disease 2019 (COVID-19) pandemic and economic impact," *Pakistan Journal of Medical Sciences*, vol. 36, no. COVID19-S4, p. S73, 2020.
- [10] S. Kumar, V. Maheshwari, J. Prabhu, M. Prasanna, P. Jayalakshmi, P. Suganya, B. A. Malar, and R. Jothikumar, "Social economic impact of COVID-19 outbreak in India," *International Journal of Pervasive Computing and Communications*, 2020.
- [11] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 115, no. 772, pp. 700–721, 1927.
- [12] F. Brauer, *Compartmental Models in Epidemiology*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 19–79.
- [13] B. Prasse and P. Van Mieghem, "Network reconstruction and prediction of epidemic outbreaks for general group-based compartmental epidemic models," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 2755–2764, 2020.
- [14] F. Fagnani and L. Zino, "Time to extinction for the SIS epidemic model: New bounds on the tail probabilities," *IEEE Transactions on Network Science and Engineering*, vol. 6, no. 1, pp. 74–81, 2017.
- [15] S. Chen, M. Small, and X. Fu, "Global stability of epidemic models with imperfect vaccination and quarantine on scale-free networks," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 3, pp. 1583–1596, 2019.
- [16] Y. Xiang, Y. Jia, L. Chen, L. Guo, B. Shu, and E. Long, "COVID-19 epidemic prediction and the impact of public health interventions: A review of COVID-19 epidemic models," *Infectious Disease Modelling*, vol. 6, pp. 324–342, 2021.
- [17] H. Rahmandad, T. Y. Lim, and J. Sterman, "Behavioral dynamics of COVID-19: Estimating underreporting, multiple waves, and adherence fatigue across 92 nations," *System Dynamics Review*, vol. 37, no. 1, pp. 5–31, 2021.
- [18] Y. Wang, "Evaluating the effectiveness of COVID-19 prevention and control measures based on SEIR model," in *Proceedings of the 2020 International Conference on Communications, Information System and Computer Engineering (CISCE)*. IEEE, 2020, pp. 143–146.
- [19] Y.-C. Chen, P.-E. Lu, C.-S. Chang, and T.-H. Liu, "A time-dependent SIR model for COVID-19 with undetectable infected persons," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 3279–3294, 2020.
- [20] C. Nowzari, V. M. Preciado, and G. J. Pappas, "Analysis and control of epidemics: A survey of spreading processes on complex networks," *IEEE Control Systems Magazine*, vol. 36, no. 1, pp. 26–46, 2016.
- [21] X. Qian and S. V. Ukkusuri, "Connecting urban transportation systems with the spread of infectious diseases: A Trans-SEIR modeling approach," *Transportation Research Part B: Methodological*, vol. 145, pp. 185–211, 2021.
- [22] M. Salehi, R. Sharma, M. Marzolla, M. Magnani, P. Siyari, and D. Montesi, "Spreading processes in multilayer networks," *IEEE Transactions on Network Science and Engineering*, vol. 2, no. 2, pp. 65–83, 2015.
- [23] B. Qu and H. Wang, "SIS epidemic spreading with heterogeneous infection rates," *IEEE Transactions on Network Science and Engineering*, vol. 4, no. 3, pp. 177–186, 2017.
- [24] H. Kang, M. Sun, Y. Yu, X. Fu, and B. Bao, "Spreading dynamics of an SEIR model with delay on scale-free networks," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 489–496, 2018.
- [25] M. Nadini, A. Rizzo, and M. Porfiri, "Epidemic spreading in temporal and adaptive networks with static backbone," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 549–561, 2018.
- [26] J. Li, T. Xiang, and L. He, "Modeling epidemic spread in transportation networks: A review," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 8, no. 2, pp. 139–152, 2021.
- [27] X.-H. Yang, B. Wang, S.-Y. Chen, and W.-L. Wang, "Epidemic dynamics behavior in some bus transport networks," *Physica A: Statistical Mechanics and its Applications*, vol. 391, no. 3, pp. 917–924, 2012.

- [28] J. Hackl and T. Dubernet, "Epidemic spreading in urban areas using agent-based transportation models," *Future Internet*, vol. 11, no. 4, p. 92, 2019.
- [29] P. E. Paré, C. L. Beck, and T. Başar, "Modeling, estimation, and analysis of epidemics over networks: An overview," *Annual Reviews in Control*, vol. 50, pp. 345–360, 2020.
- [30] M. Ye, J. Liu, C. Cenedese, Z. Sun, and M. Cao, "A network SIS meta-population model with transportation flow," in *Proceedings of the IFAC World Congress*, 2020, pp. 2562–2567.
- [31] Q. Yin, Z. Wang, C. Xia, M. Dehmer, F. Emmert-Streib, and Z. Jin, "A novel epidemic model considering demographics and intercity commuting on complex dynamical networks," *Applied Mathematics and Computation*, vol. 386, p. 125517, 2020.
- [32] D. Brockmann and D. Helbing, "The hidden geometry of complex, network-driven contagion phenomena," *Science*, vol. 342, no. 6164, pp. 1337–1342, 2013.
- [33] F. Iannelli, A. Koher, D. Brockmann, P. Hövel, and I. M. Sokolov, "Effective distances for epidemics spreading on complex networks," *Physical Review E*, vol. 95, no. 1, p. 012313, 2017.
- [34] L. Zhong, M. Digne, W. Wang, and J. Gao, "Country distancing increase reveals the effectiveness of travel restrictions in stopping COVID-19 transmission," *Communications Physics*, vol. 4, no. 1, p. 121, 2021.
- [35] Anupriya, P. Bansal, and D. J. Graham, "Modelling the propagation of infectious disease via transportation networks," *Scientific Reports*, vol. 12, no. 1, p. 20572, 2022.
- [36] S. G. Yücel, R. H. Pereira, P. S. Peixoto, and C. Q. Camargo, "Impact of network centrality and income on slowing infection spread after outbreaks," *Applied Network Science*, vol. 8, no. 1, p. 16, 2023.
- [37] P. Di Giamberardino, D. Iacoviello, F. Papa, and C. Sinisgalli, "A data-driven model of the COVID-19 spread among interconnected populations: Epidemiological and mobility aspects following the lockdown in Italy," *Nonlinear Dynamics*, vol. 106, pp. 1239–1266, 2021.
- [38] J. Arino and P. van den Driessche, "A multi-city epidemic model," *Mathematical Population Studies*, vol. 10, no. 3, pp. 175–193, 2003.
- [39] X. Liu and P. Stechlin, "Transmission dynamics of a switched multi-city model with transport-related infections," *Nonlinear Analysis: Real World Applications*, vol. 14, no. 1, pp. 264–279, 2013.
- [40] D. Knipf, "Stability criteria for a multi-city epidemic model with travel delays and infection during travel," *Electronic Journal of Qualitative Theory of Differential Equations*, vol. 2016, no. 74, pp. 1–22, 2016.
- [41] B. S. Pujari and S. Shekatkar, "Multi-city modeling of epidemics using spatial networks: Application to 2019-nCoV (COVID-19) coronavirus in India," *Medrxiv*, pp. 2020–03, 2020.
- [42] J. Arino and P. Van den Driessche, "Disease spread in metapopulations," *Fields Institute Communications*, vol. 48, no. 1, pp. 1–13, 2006.
- [43] L. Wang and X. Li, "Spatial epidemiology of networked metapopulation: An overview," *Chinese Science Bulletin*, vol. 59, no. 28, pp. 3511–3522, 2014.
- [44] Q. Shao and D. Han, "Epidemic spreading in metapopulation networks with heterogeneous mobility rates," *Applied Mathematics and Computation*, vol. 412, p. 126559, 2022.
- [45] X. Zhu, Y. Liu, S. Wang, R. Wang, X. Chen, and W. Wang, "Allocating resources for epidemic spreading on metapopulation networks," *Applied Mathematics and Computation*, vol. 411, p. 126531, 2021.
- [46] D. Vrabac, M. Shang, B. Butler, J. Pham, R. Stern, and P. E. Paré, "Capturing the effects of transportation on the spread of COVID-19 with a multi-networked SEIR model," *IEEE Control Systems Letters*, vol. 6, pp. 103–108, 2022.
- [47] B. Butler, C. Zhang, I. Walter, N. Nair, R. Stern, and P. E. Paré, "The effect of population flow on epidemic spread: Analysis and control," in *Proceedings of the 2021 60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 4260–4265.
- [48] R. S. Varga, *Iterative Analysis*. Springer, 1962.
- [49] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 2012.
- [50] H. Shu, D. Fan, and J. Wei, "Global stability of multi-group SEIR epidemic models with distributed delays and nonlinear transmission," *Nonlinear Analysis: Real World Applications*, vol. 13, no. 4, pp. 1581–1592, 2012.
- [51] M. Hu et al., "Risk of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Transmission Among Air Passengers in China," *Clinical Infectious Diseases*, vol. 75, no. 1, pp. e234–e240, 09 2021.
- [52] B. A. Butler, "Networked SEIRS with travel flows," https://github.com/brooksbutler/SEIRS_PopulationFlows, 2023.
- [53] "Minnesota coronavirus cases and deaths," [https://usafacts.org/visualizations/coronavirus-\(COVID-19\)-spread-map/state/minnesota](https://usafacts.org/visualizations/coronavirus-(COVID-19)-spread-map/state/minnesota), Accessed September 16, 2022.
- [54] "Street Light Data," <https://www.streetlightdata.com/>, Accessed November 2, 2022.



Brooks Butler (Member, IEEE) is a Postdoctoral Fellow at the University of California, Irvine, where he is a participant in the Intelligence Community Postdoctoral Research Fellowship Program administered by Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the U.S. Department of Energy and the Office of the Director of National Intelligence (ODNI). He received his Ph.D. in Electrical and Computer Engineering from Purdue University in 2024. He received his M.S. in Computer Science and his B.S. in Applied Physics from Brigham Young University in 2020 and 2019, respectively.



Raphael Stern (Member, IEEE) is an Assistant Professor with the Department of Civil, Environmental, and Geo- Engineering at the University of Minnesota. He received the B.Sc., M.Sc., and Ph.D. degrees in civil engineering from the University of Illinois at Urbana-Champaign in 2013, 2015, and 2018, respectively. His research interests include transportation data analysis and traffic flow modelling.



Philip E. Paré (Senior Member, IEEE) is the Rita Lane and Norma Fries Assistant Professor of Electrical and Computer Engineering at Purdue University. He received his Ph.D. in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 2018, after which he went to KTH Royal Institute of Technology in Stockholm, Sweden to be a Post-Doctoral Scholar. He received his B.S. in Mathematics with University Honors and his M.S. in Computer Science from Brigham Young University in 2012 and 2014, respectively. He is a 2023 recipient of the NSF CAREER award, a 2021 Societal Impact Fellow (Purdue), and a 2023 Teaching for Tomorrow Fellow (Purdue). His research focuses on networked control systems, namely modeling, analysis, and control of virus spread over networks.