

Estimating epidemic arrival times using linear spreading theory

Lawrence M. Chen,¹ Matt Holzer,² and Anne Shapiro³

¹*Department of Mathematics, University of Kansas, Lawrence, KS 66045, USA*

²*Department of Mathematical Sciences, George Mason University, Fairfax, VA 22030, USA*

³*Department of Mathematics and Statistics, Carleton College, Northfield, MN 55057, USA*

(Dated: August 16, 2017)

We study the dynamics of a spatially structured model of worldwide epidemics and formulate predictions for arrival times of the disease at any city in the network. The model is comprised of a system of ordinary differential equations describing a meta-population SIR compartmental model defined on a network where each node represents a city and edges represent flight paths connecting cities. Making use of the linear determinacy of the system, we consider spreading speeds and arrival times in the system linearized about the unstable disease free state. Two predictions are presented. The first assumes that the dominant transmission pathway between any two cities can be approximated by a one dimensional lattice or homogeneous tree and gives a uniform prediction for arrival times independent of specific network features. The second method considers the full heat kernel solution of the linear system and produces an arrival time estimate that incorporates network features. We test these predictions on a real network describing worldwide airline traffic.

I. INTRODUCTION

We study the spread of an infectious disease through a global airline network. Among the fundamental properties of interest are the arrival times of the disease. Given a disease that originates in a particular city or location, how long does it take for that disease to arise in some other city? In this paper, we develop methods to estimate arrival times based upon the theory of traveling fronts.

Mathematical modeling of infectious diseases often involves compartmental models where the total population is divided into groups pertaining to a certain state, i.e. susceptible, infected, recovered, etc.. These populations interact and their evolution is described by a system of differential equations, see [9, 20]. These models can be extended to include spatial structure which often leads to partial differential equation analogs of the local models. Traditional analysis of the spatial spread of disease has involved the study of traveling fronts; see [5, 17]. Ostensibly, the prevalence of long range connections between cities via airline travel would render this approach ineffective. On the contrary, we demonstrate in this article that when the rate of diffusion is small relative to the reaction rates then the dynamics are still dominated by the formation of traveling fronts only with a different notion of distance between cities to accommodate the long range connections; see [3, 19].

The model considered here is an SIR (susceptible-infected-recovered) meta-population model; see [24]. This model has been employed to great effect to simulate worldwide epidemics; see for example [6, 7, 12, 18]. In this model, the world is described by a graph whose nodes correspond to cities and whose edges correspond to flights paths connecting cities. With local reaction terms describing the disease dynamics and diffusive transport between nodes, the model is a reaction-diffusion equation on a network. The study of dynamical systems on networks is a burgeoning field and we refer the reader to

[1, 21–23] and the references therein.

Our goal in this article is to estimate arrival times of an epidemic spreading through a worldwide airline network. We make use of the linear determinacy of the system; see [5, 17], and apply some tools from linear spreading theory; see [25], to formulate estimates for the system linearized about the unstable disease free state. We compare these predictions to observed arrival times in the nonlinear system determined from numerical simulations of the disease spread.

To summarize, we provide two predictions for arrival times. The first approach is to approximate the pathway connecting any two cities by a one dimensional lattice or homogeneous tree and then compute the linear spreading speed of solutions propagating along that corridor. The second approach builds on the first and incorporates additional information regarding the structure of the network by working with approximations of the full heat kernel for the linearization about the unstable state. The linear spreading speed method (LSS) requires no information about the network aside from distances between cities, i.e. the minimal number of flights required to transport from one city to the other. The heat kernel (HK) approach is more accurate as a method of estimation, but requires knowledge of the weights present in the adjacency matrix. We show that as the diffusion parameter tends to zero, the two estimates agree at first and second order.

Previous authors have also considered disease propagation in complex networks through the perspective of traveling fronts; see for example [2, 14, 18]. Our approach is partially motivated by recent work on invasion fronts for the Fisher-KPP equation in random networks; see [15]. We note previous work related to estimating epidemic arrival times; see [3, 10, 11, 19]. These estimates are primarily based upon probabilistic considerations and are related, but distinct, from the methods developed here and our efforts complement their approaches.

The paper is organized as follows. In section II, we

review the meta-population SIR model and worldwide airline network that will be the primary object of study. In section III, we develop two methods to predict arrival times. In section IV, we apply our methods to estimate arrival times in numerical simulations of the model. Finally, we conclude in V with a discussion.

II. META-POPULATION SIR MODEL

We study the following meta-population susceptible-infected-reduced (SIR) model described in [3]:

$$\begin{aligned}\partial_t s_n &= -\alpha s_n j_n + \gamma \sum_{m \neq n} P_{nm} (s_m - s_n) \\ \partial_t j_n &= \alpha s_n j_n - \beta j_n + \gamma \sum_{m \neq n} P_{nm} (j_m - j_n) \\ \partial_t r_n &= \beta j_n + \gamma \sum_{m \neq n} P_{nm} (r_m - r_n).\end{aligned}\quad (1)$$

Here s_n , j_n and r_n are the normalized susceptible, infected and recovered populations of city n , respectively. The parameter α describes the rate of infection and β is the mean recovery rate of individuals. The parameter γ is the average mobility rate, meaning the percentage of the population of the system traveling at any given time. These parameters govern the spread of the disease. Expressing the dependent variables in vector notation, we see that the disease free state $(\mathbf{s}, \mathbf{j}) = (\mathbf{1}, \mathbf{0})$ is an equilibrium solution and is unstable so long as $\alpha > \beta$. As such, if the disease free state is perturbed due to the introduction of some infected population at one node then the infected population will grow in number and spread into the surrounding network.

Mathematically, (1) is a system of reaction-diffusion equations on a graph $G = (V, E)$. The nodes correspond to cities and edges represent the presence of air traffic connecting the two cities. The matrix \mathbf{P} is a weighted adjacency matrix with $0 \leq P_{nm} \leq 1$. It is row stochastic, with each P_{nm} corresponding to the probability of a random walker at node n moving to node m . Due to the row stochasticity of \mathbf{P} , observe that total population at each city is preserved and we may recover r_n using $r_n = 1 - s_n - j_n$.

Since we are interested in epidemics spreading via an airline transportation network, we work with the World Airline Network (WAN) provided by [13]. This network is generated from flights scheduled on 800 airlines for the period November 1, 2000, to October 31, 2001. It includes all scheduled flights and charter flights, both for large (air carriers) and small aircrafts (air taxis). Although the data is now 16 years old, it can still be considered reflective of the current WAN. Note that cities are considered as opposed to airports and therefore airports such as O'Hare and Midway are grouped together and considered as the single city Chicago.

The WAN consists of 3618 cities (nodes). There are 14,142 connections (edges). The diameter of the graph

is 17, representing the seventeen total flights required to transport from Royal Air Force Brize Norton in Oxfordshire, England, to Wasu, Papua New Guinea. A number of graph theoretic properties of the network are studied in [13]. The network shares many properties consistent with scale-free random graphs such as Barabasi-Albert [8, 26].

Since the data only includes connections between airports, we need to specify the matrix \mathbf{P} that assigns weights to various connections within the network. Several authors have developed quantitative estimates for these weights based upon city populations and air traffic data; see for example [3]. We proceed more abstractly and will concentrate on three different types of weights:

- The *uniform* \mathbf{P} matrix is meant to simulate a simple random walk along the network. As such, in each row every non-zero entry is identical since a walker starting in node n has no preference for any of its adjacent nodes m . Here, $\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$, where \mathbf{D} is the diagonal degree matrix of the network, and \mathbf{A} is the unweighted, undirected adjacency matrix.
- The *proportional* \mathbf{P} matrix supposes that a walker may prefer to travel towards nodes with higher connectivity, i.e. local hubs. To implement this, each non-zero entry in each row of \mathbf{A} was assigned a value equal to the column node's degree. The rows were normalized to maintain the matrix's stochastic property.
- A *randomized* \mathbf{P} matrix randomly assigns weights to each edge in the network. This was achieved by assigning each non-zero entry in each row of \mathbf{A} a random integer value between 1 and 100, inclusive, and then the rows were normalized.

We now turn our attention to arrival times.

III. ARRIVAL TIMES: PREDICTIONS

Our primary interest in this article is to study how a theoretical epidemic will spread through the network. To this end, we focus on initial conditions where $j_n(0) = j_0$ for some city n and $j_k(0) = 0$ for all other cities. In Figure 1, we plot the infected proportion at three cities (Paris, Kodiak and Wasu) given an epidemic originating in Paris.

We will focus on arrival times, meaning the first time at which a disease is established in a certain city. Mathematically, for a disease initialized in city n we define the arrival time at city m as

$$T_{nm} = \inf\{ t \geq 0 \mid j_m(t) = \kappa \},$$

for some threshold κ . Note that T_{nm} depends on the specific cities under consideration, system parameters, the initial infection proportion j_0 and the threshold κ . We

suppress this functional dependence, but will comment on the dependence on j_0 and κ later.

Central to our approach is the *linear determinacy* of the SIR model (1). Linear determinacy generally applies to reaction-diffusion PDEs models and refers to the property that the speed of invasion fronts in that context can be determined from the linear spreading speed of perturbations in the linearization about the unstable state. See [17] for the PDE case and [5] for recent work on a one dimensional lattice.

In the context of (1), we can again linearize the system about the disease free steady state $(\mathbf{s}, \mathbf{j}) = (\mathbf{1}, \mathbf{0})$ and obtain

$$\begin{aligned}\partial_t \mathbf{s}_l &= -\alpha \mathbf{j}_l + \gamma(\mathbf{P} - \mathbf{I})\mathbf{s}_l \\ \partial_t \mathbf{j}_l &= (\alpha - \beta)\mathbf{j}_l + \gamma(\mathbf{P} - \mathbf{I})\mathbf{j}_l.\end{aligned}\quad (2)$$

Note that the equation for \mathbf{j}_l decouples and an explicit solution for this linear system can be written as

$$\mathbf{j}_l(t) = j_0 e^{\Gamma t} e^{\gamma P t} v_n,$$

where we have introduced $\Gamma = \alpha - \beta - \gamma$ and v_n is the standard Euclidean basis vector. Denote \mathbf{j} as the solution of the nonlinear system (1). By a comparison principle argument we can show that $\mathbf{j}_l(t) \geq \mathbf{j}(t)$ for all $t > 0$, assuming identical initial conditions. To see this, define $Z = \mathbf{j}_l - \mathbf{j}$. The differential equation for the difference $Z(t)$ is,

$$\frac{dZ}{dt} = -\beta Z + \gamma(\mathbf{P} - \mathbf{I})Z + \alpha(\mathbf{j}_l - \mathbf{s}\mathbf{j}),$$

where the product $\mathbf{s}\mathbf{j}$ is understood to be component-wise multiplication. Assume that one component of Z were to become negative. This is impossible, since when that component was zero we would necessarily have that $\frac{dZ}{dt} > 0$ due to the positivity of Z at nearby nodes and the component wise bound $\mathbf{s} \leq 1$.

This leads to the important observation that the arrival time of the linear problem always provides a lower bound on the arrival time in the nonlinear problem. We therefore concern ourselves with determining the arrival times for the linear system,

$$T_{nm} = \inf\{t \geq 0 \mid j_0 v_m^T e^{\Gamma t} e^{\gamma P t} v_n = \kappa\}.$$

We now proceed to estimate the arrival times in the linear problem.

A. Linear Spreading Speed

We now formulate a prediction for the arrival time based upon linear spreading theory. The key assumption here is that the shortest path(s) – in terms of flights required – connecting the cities m and n constitute the dominant transmission pathway of the disease between those cities. This corridor is then assumed to be well

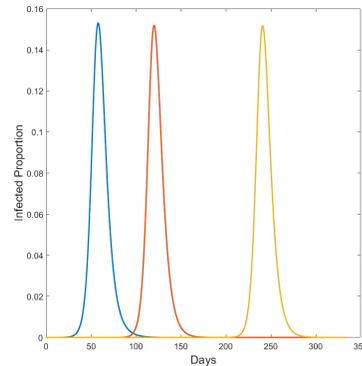


Figure 1: Proportion of infected population in Paris, France (blue), Kodiak, Alaska (Red), and Wasu, Papua New Guinea (Yellow) as a function of time with infection starting in Paris. Simulation was ran using uniform P matrix. Note that in all three cities, the infected population reaches the nearly the same proportion. $\gamma = 0.001$

approximated by a one dimensional lattice or a homogeneous tree. For the one dimensional lattice this assumption translates to one where the loss of infected population to the remaining network is either negligible or balanced by transport in the opposite direction. For the homogeneous tree, the assumption is that the loss of infected population to the rest of the network can be suitably approximated by a tree.

Under this assumption the dynamics of (2) along the shortest path are approximated by the following linear, lattice dynamical system

$$\partial_t j_n = (\alpha - \beta)j_n + \frac{\gamma}{k+1}[j_{n-1} - (k+1)j_n + k j_{n+1}]. \quad (3)$$

Here $k = 1$ for a one dimensional lattice and $k > 1$ for a homogeneous tree with branching factor k . To determine the linear spreading speed of solutions of (3), we first substitute the ansatz $j_n(t) = e^{\lambda t - \nu n}$, from which the dispersion relation,

$$\lambda = (\alpha - \beta) + \frac{\gamma}{k+1}[e^{\nu} - (k+1) + k e^{-\nu}] \quad (4)$$

relates solutions with spatial decay rate ν to their corresponding temporal growth rate λ . With $\lambda(\nu)$ in hand, the envelope speed of $e^{-\nu n}$ is

$$s_{env}(\nu) = \frac{\lambda(\nu)}{\nu}. \quad (5)$$

The envelope speed gives the speed at which an exponentially decaying solution $e^{-\nu n}$ spreads in the linear system. Minimizing over all such ν we obtain the *linear spreading speed*,

$$s_{lin} = \min_{\nu > 0} s_{env}(\nu).$$

Differentiating (5) with respect to ν , we see that critical points can be found by simultaneously solving the system of equations

$$s = \frac{\lambda(\nu)}{\nu}, \quad s = \partial_\nu \lambda(\nu). \quad (6)$$

Roots of this system of equations are known as pinched double roots; see [16, 25]. Setting $\frac{\lambda(\nu)}{\nu} = \partial_\nu \lambda(\nu)$, we obtain the following equation,

$$0 = (\alpha - \beta) + \frac{\gamma}{k+1} [(1 - \nu)e^\nu - (k+1) + (1 + \nu)ke^{-\nu}]. \quad (7)$$

We write $\nu_{lin}(\gamma)$ as the solution of (7), from which

$$s_{lin}(\gamma) = \frac{\gamma}{k+1} \left(e^{\nu_{lin}(\gamma)} - ke^{-\nu_{lin}(\gamma)} \right). \quad (8)$$

In general, we do not have analytical formulas for the linear spreading speed and speeds must be computed numerically. The linear spreading speed (LSS) approximation provides one spreading speed throughout the entire network for any given set of initial parameters. The quantity $\frac{1}{s_{lin}(\gamma)}$ then provides an estimate for the inter-city arrival times in the network. Multiplying by the minimal number of flights, d , required to traverse the network from city n to city m , we can obtain a prediction for the arrival times, which we denote

$$T_{nm}^{LSS} = \frac{d}{s_{lin}(\gamma)} \quad (9)$$

The arrival time estimate T_{nm}^{LSS} does not incorporate any specific information concerning the topology or weights of the transportation network. This is both a strength and a weakness. On the positive side, this prediction requires very little information to formulate a prediction for arrival times. Of course, on the other hand, we expect the accuracy of this prediction to suffer due to the limited details of the network incorporated. In the next section, we formulate a second prediction based upon the full heat kernel of the graph.

B. Heat Kernel

We now refine our arrival time prediction to include more information about the network. To do this, we consider the full linear evolution of the infected population in (2),

$$j_m(t) = j_0 v_m^T e^{\Gamma t} e^{\gamma P t} v_n,$$

where again we have assumed that the infection begins at city n with initial infected proportion j_0 . We expand the matrix exponential $e^{\gamma P t}$ as a series

$$v_m^T e^{\gamma P t} v_n = \sum_{k=0}^{\infty} \frac{\gamma^k t^k}{k!} v_m^T P^k v_n. \quad (10)$$

Let $\rho_k = v_m^T P^k v_n$. Note that ρ_k describes the probability of a random walker starting at city m being located in city n after k steps. Suppose that the minimal number of flights required to travel from city n to city m is d . Then, in the expansion

$$j_m(t) = j_0 e^{\Gamma t} \left[\rho_0 + \gamma \rho_1 t + \dots + \frac{\gamma^d \rho_d t^d}{d!} + \dots \right], \quad (11)$$

all terms before the d th term must be zero.

To formulate an arrival time prediction based upon the heat kernel expansion, we assume that the leading order term in the expansion dominates for small γ . That is, the dominant contribution to the disease spreading from node n to m is along the shortest path(s) of length d .

By solving,

$$j_0 e^{\Gamma t} \frac{\gamma^d \rho_d t^d}{d!} = \kappa,$$

we obtain the heat kernel estimate for the arrival time

$$T_{nm}^{HK} = \frac{d}{\alpha - \beta - \gamma} W \left(\frac{(d!)^{1/d}}{d} \frac{\alpha - \beta - \gamma}{\gamma (\rho_d)^{1/d}} \left(\frac{\kappa}{j_0} \right)^{1/d} \right), \quad (12)$$

where W is once again the Lambert-W function.

In contrast to T_{nm}^{LSS} , the heat kernel estimate includes information regarding the number of shortest paths as well as the weights of the connections along those paths. Thus, we expect this estimate to be more accurate.

Validity of the heat kernel prediction depends on the leading order term in (10) dominating. There are two general ways in which this assumption can fail. The first involves cases where $\rho_d \ll \rho_{d+1}$, suggesting that the second term in the heat kernel expansion can not be ignored. A second case occurs for larger values of γ , where accurately estimating the sum (10) requires multiple terms. Therefore, we expect the heat kernel estimate to closely approximate the linear dynamics in the asymptotic limit as $\gamma \rightarrow 0$. If γ is no longer small, then we see that the shortest paths need not be the most dominant contributor to disease spread along the network. This deficiency can be remedied by including more terms in the the heat kernel. For example, we could consider

$$e^{\Gamma t} \left[\frac{\gamma^d \rho_d t^d}{d!} \left(1 + \frac{\gamma t}{(d+1)} \frac{\rho_{d+1}}{\rho_d} + \dots \right) \right] = \frac{\kappa}{j_0}. \quad (13)$$

Instead of neglecting the terms after the d th term, the estimate can be refined to include as many terms as needed.

C. Comparison of predictions

We now compare the two estimates T_{nm}^{LSS} and T_{nm}^{HK} . For simplicity, we set $j_0 = \kappa$; see the discussion in the second paragraph of Section IV.

Recall that we are interested in the limit $\gamma \rightarrow 0$ and focus on that case here. We first obtain expansions for

T_{nm}^{LSS} in this limit. Recall (7). Assuming that $\alpha - \beta$ is positive and independent of γ , we then require that $\nu \rightarrow \infty$ to balance the asymptotically small γ terms. In doing so, we must solve

$$\alpha - \beta = \frac{\gamma}{k+1}(\nu - 1)e^\nu.$$

This yields an estimate for the linearly selected decay rate

$$\nu_{lin} \approx W\left(\frac{(\alpha - \beta)(k+1)}{e\gamma}\right) + 1,$$

where W is the Lambert-W function. Since the arrival time is approximately $\frac{d}{s_{lin}}$, we see from (8) that

$$T_{nm}^{LSS} \approx \frac{d}{\alpha - \beta} W\left(\frac{(\alpha - \beta)(k+1)}{e\gamma}\right) \quad (14)$$

Asymptotic expansions of both expressions are readily available using properties of the Lambert W function. We obtain,

$$T_{nm}^{LSS} = \frac{-d}{\alpha - \beta} \log \gamma - \frac{d}{\alpha - \beta} \log(-\log \gamma) + \frac{d}{\alpha - \beta} \log\left(\frac{(\alpha - \beta)(k+1)}{e}\right) + o(1), \quad (15)$$

and

$$T_{nm}^{HK} = \frac{-d}{\alpha - \beta} \log \gamma - \frac{d}{\alpha - \beta} \log(-\log \gamma) - \frac{d}{\alpha - \beta} \log \frac{d}{\alpha - \beta} - \frac{1}{\alpha - \beta} \log \frac{\rho_d}{d!} + o(1). \quad (16)$$

As $\gamma \rightarrow 0$ arrival times tend to infinity. Both arrival time estimates agree to first and second order in their expansions. They differ at $\mathcal{O}(1)$. At this juncture, we compare our arrival time estimates to those of [11]. In fact, three different arrival time estimates are provided in [11]: referred to as front, Gumbel and deterministic. The deterministic formulation involves the Lambert-W function and is the most related to the heat-kernel and linear spreading speed estimates provided here. In the limit as $\gamma \rightarrow 0$, the deterministic, heat kernel and linear spreading speed estimates are equivalent to $\mathcal{O}(\log(-\log(\gamma)))$. The Gumbel estimate is equivalent to T_{nm}^{LSS} and T_{nm}^{HK} to leading order, but has no $\log(-\log(\gamma))$ correction.

Note that the dependence of T_{nm}^{LSS} on the branching factor k does not appear until $\mathcal{O}(1)$. In the numerical simulations in section IV, we typically take k to be one less than the average degree in the network. For the WAN from [13], this is $k = 6.871$. The optimal choice of k is not obvious, but we find that this choice gives reasonable results.

Comparing expressions (14) and (12) directly, we could match the arguments of the Lambert W functions

$$\frac{\alpha - \beta}{e\gamma}(k+1) = \frac{(d!)^{1/d}}{d} \frac{\alpha - \beta}{\gamma(\rho_d)^{1/d}}.$$

Note that for large d

$$\frac{(d!)^{1/d}}{d} \approx e^{-1}.$$

This suggests that k should be chosen according to

$$k+1 = \frac{1}{(\rho_d)^{1/d}}.$$

Thus, if there was a single shortest path connecting two nodes, the tree with degree $k+1$ given by the geometric mean of the weights connecting the two cities is a likely choice. Since typical shortest paths pass through high degree hubs, this would suggest large values of k in general. That being said, it is rarely the case that there is a single shortest path connecting two cities. Thus, for simplicity, we take k to be one less than the mean degree.

IV. ARRIVAL TIMES: PREDICTIONS VERSUS NUMERICAL OBSERVATIONS

We now apply our predictions to system (1). For the purposes of simulations, we fix the parameters $\alpha = 0.5$ and $\beta = 0.25$. We vary γ between 0.001 and 0.1 and note that this range includes the estimated range for airline traffic derived in [3]. Numerical simulations are performed using explicit Euler with typical timestep $\Delta t = 0.01$.

A remark regarding the initial condition j_0 and threshold κ is in order. As a general rule, the smaller these quantities are the better the estimates are. In simulations, we have typically set $j_0 = \kappa$ and used values between 0.01 (arrival when one percent of the population is infected) and 10^{-6} (arrival when approximately one individual is infected). This range of values also give satisfactory results even if $j_0 \neq \kappa$. Too large of values for j_0 can be problematic since the assumption is that the system is well approximated by its linearization, at least on the portion of the network where the interesting dynamics are occurring, i.e. ahead of the front interface. If one insists on larger values of j_0 , then there is typically some initial transient before the estimates are valid. For the linear spreading speed estimate this may not be an issue, however in the heat kernel estimate this can mean that the estimates are off by some fixed constant.

The heat kernel estimate gives a prediction for each pair of cities, n and m . In contrast, the linear spreading speed prediction gives one prediction for the entire network. Therefore, we will compare the linear spreading speed to the average speed of infection in the network as follows. We first separate cities into groups according their distance from the initial infection location. We then compute the mean arrival time at each distance and compute a linear regression line with these data points. We exclude both the original infection location and all cities in Papa New Guinea in these calculation. For this version of the WAN, Papa New Guinea contains a sequence

of cities that are the furthest distance from a majority of cities in the network; see those cities in Figure 3 with distance nine and greater from the origin city. Including them in the averaged arrival times gives undue importance to these nodes, so we exclude them here. The reciprocal of the slope then gives as estimate for the average time required for the disease to spread between adjacent nodes on the graph. We claim that $\frac{1}{s_{lin}}$ provides a reasonable leading order approximation.

A. Uniform P

We first consider the case of $P = D^{-1}A$ describing a random walk on the network where weights at each node are chosen evenly. In Figure 2, we compare averaged inter-city invasion times to the estimate from the linear spreading speed. We randomly select one hundred cities in which to initialize the disease and compute average invasion times. Comparing two different values of γ , we find that the linear spreading speed is a reasonable leading order estimate.

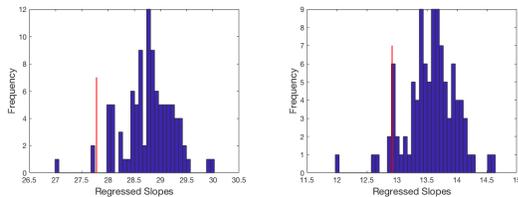


Figure 2: Histogram of regressed slopes taken from 100 random cities from using uniform $P = D^{-1}A$. On the left is $\gamma = 0.0001$ and on the right is $\gamma = 0.01$. The red line denotes the prediction based upon the linear spreading speed.

To focus our discussion, we will consider infections starting in only two locales: Paris, France and Kodiak, USA. We select these nodes since Paris is the city with the largest number of flights and Kodiak is a periphery node with only one connection to Anchorage. Arrival times as a function of distance from the initial node are shown in Figure 3 for Kodiak and in Figure 4 for Paris. Note the general linear relationship exhibited by the data with nodes further from the source taking longer for the infection to arrive at these nodes than closer nodes. Simultaneously, we note that there remains significant deviations in the arrival times of the infection even among nodes that are a fixed distance from the source node. We compile comparisons between arrival times observed in numerical simulation to linear spreading speed for infection starting in Paris and Kodiak over varying γ together with relative errors (REs) in Table I. All units are in days. Again, regression times were calculated disregarding the origin city and the cities in Papa New Guinea.

The linear spreading speed provides a reasonable leading order estimate for arrival times, but significant deviations remain between this speed and observed speeds. For example, with $\gamma = 0.01$ and an infection starting in Paris

γ	LSS	Kodiak RE	Kodiak	Paris RE	Paris
0.001	19.961	20.779	0.0394	19.223	0.0383
0.01	12.920	13.435	0.0383	11.950	0.0811
0.1	8.036	7.636	0.0522	6.281	0.2791

Table I: Arrival time per flight calculated from the linear spreading speed and from numerical observations for a range of γ values and epidemics originating in Kodiak and Alaska.

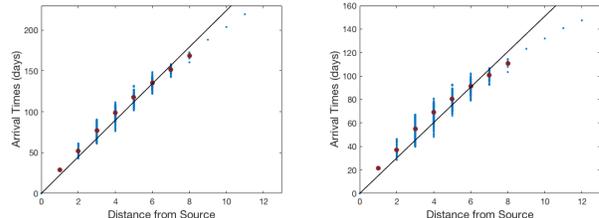


Figure 3: Arrival times compared to distance from infection source node with $\gamma = 0.001$ and $\gamma = 0.01$ and infection originating in Kodiak, Alaska. Red dots represents mean arrival time at each distance. The ultimate 4 nodes represent cities in Papua New Guinea and were not included in analysis. For $\gamma = 0.001$, the regressed slope gives arrival times of 20.779 between cities while the linear spreading speed predicts arrival times of 19.961 days. For $\gamma = 0.01$, the regressed slope gives arrival times of 13.435 between cities while the linear spreading speed predicts arrival times of 12.920 days.

those cities that are three flights away from the source exhibit arrival time variation exceeding twenty five days. We expect the heat kernel arrival time to do better, since it incorporates more information about the transmission paths between any two cities.

The heat kernel method is indeed more accurate. In Figure 5, predicted arrival times versus observed arrival times for an infection originating in Paris are plotted and we find excellent agreement between the two with mean error absolute errors of approximately 0.40 days. Correspondence is even better for smaller values of γ . Similar results hold for infections originating in Kodiak, see Figure 6.

For γ values larger than 0.01, we find that the heat kernel approximation also exhibits errors. We work with a specific example of an infection originating in Kodiak and arriving at Yellowknife, Canada. The shortest path between these two cities consists of four flights and involves first flying to Anchorage, followed by one of five hubs (Chicago, Los Angeles, Minneapolis-St. Paul, Seattle, Toronto), followed by flights to Edmonton and then Yellowknife. For asymptotically small values of γ this is the dominant transmission path. However, for larger values of γ there are longer paths with non-negligible contribution to the arrival times. For example, one can fly from Anchorage to Juneau to Whitehorse to Fort Simpson to Yellowknife. This route requires five flights, but the number of flights emanating from these cities is very

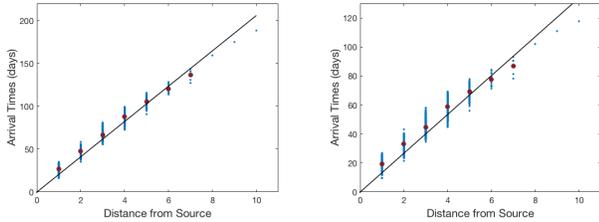


Figure 4: Arrival times compared to distance from infection source node with $\gamma = 0.001$ and infection originating in Paris, France. Red dots represents mean arrival time at each difference. The ultimate 3 nodes represent cities in Papua New Guinea and were not included in analysis. For $\gamma = 0.001$, the regressed slope gives arrival times of 19.223 between cities while the linear spreading speed predicts arrival times of 19.961 days. For $\gamma = 0.01$, the regressed slope gives arrival times of 11.950 between cities while the linear spreading speed predicts arrival times of 12.920 days.

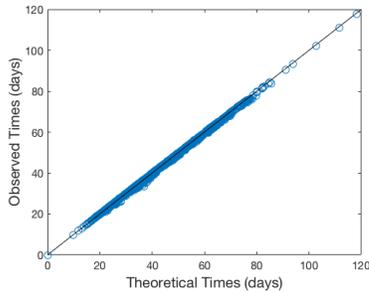


Figure 5: Simulated arrival times compared to one-term heat kernel approximation using uniform P matrix for infection originating in Paris, France with $\gamma = 0.01$ plotted against the line $y = x$.

low compared to the shorter path. This leads to $\frac{\rho_5}{\rho_4} \approx 8$ for this example, which translates to saying that it is eight times more likely for a random walker arriving in Yellowknife to have been in Kodiak five steps before as opposed to four. Therefore, when γ is larger the most probable path is not necessarily the shortest.

A second issue is that when γ is not asymptotically small, one may require more terms in the heat kernel expansion to get a reasonable approximation. For example, supposing that the arrival time scales with $\log \gamma$, then the terms $(\gamma t)^k$ do not converge rapidly to zero and more than one term is likely required to get an accurate approximation. In Figure 8, we plot arrival times in Yellowknife for an infection originating in Kodiak. The one term heat kernel exhibits errors due to the mechanism just described, but the two term heat kernel closely matches the full heat kernel estimate over a large range of γ values.

In Figure 7, one observes the errors that are present in the one term heat kernel expansion for an epidemic initialized in Kodiak. Here $\gamma = 0.1$ is unrealistically large for the application at hand, but we include it for illustra-

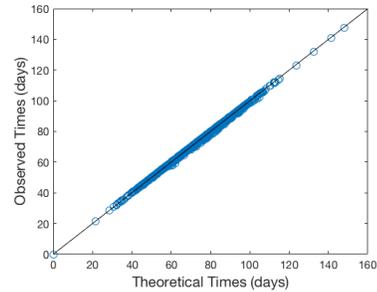


Figure 6: Simulated arrival times compared to one-term heat kernel approximation using uniform P matrix for infection originating in Kodiak, Alaska with $\gamma = 0.01$ plotted against the line $y = x$.

tive purposes. We see that including more terms in the heat kernel expansion leads to a more accurate result; see Figure 9.

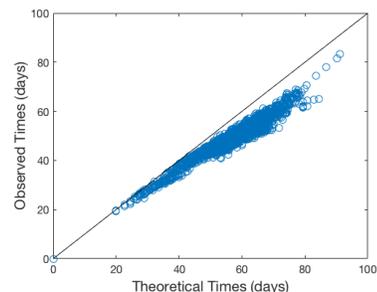


Figure 7: Simulated arrival times compared to one-term heat kernel approximation using uniform P matrix for infection originating in Kodiak, Alaska with $\gamma = 0.1$ plotted against the line $y = x$.

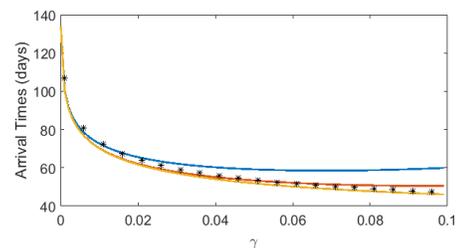


Figure 8: Arrival times in Yellowknife as a function of γ for an infection starting in Kodiak (asterisks). In blue is the one term heat kernel approximation, orange is the two term heat kernel and the fifteen term heat kernel is in yellow.

B. Proportional and Random P

A major challenge in estimating arrival times, even when γ is very small, is attaining precise information regarding the weights in the matrix P. It is often the

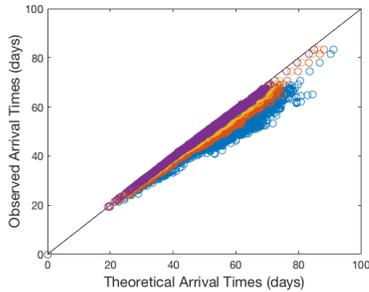


Figure 9: Simulated arrival times compared to one, two, three, and four-term heat kernel approximations using uniform P matrix for infection originating in Kodiak, Alaska with $\gamma = 0.1$ plotted against the line $y = x$. One-term heat kernel is plotted in blue, two term is plotted in orange, three term is plotted in yellow, and four term is plotted in purple. As more terms are added to the heat kernel approximation, it grows closer to the line.

case that the edges of the network can be determined, but that estimating fluxes and the corresponding weights is more difficult.

In light of the success of the heat kernel in estimating arrival times for small γ , we must comment on the utility of the linear spreading speed estimate. A major advantage of the linear spreading speed estimate is that it does not require information regarding the topology or weights of a network, but only knowledge of the distance of the shortest paths connecting any two nodes. In Figure 10 we compile average invasion times found via linear regression and compare with the linear spreading speed. In both cases, the linear spreading speed provides a reasonable leading order approximation without any knowledge of the weights in the network aside from the average degree.

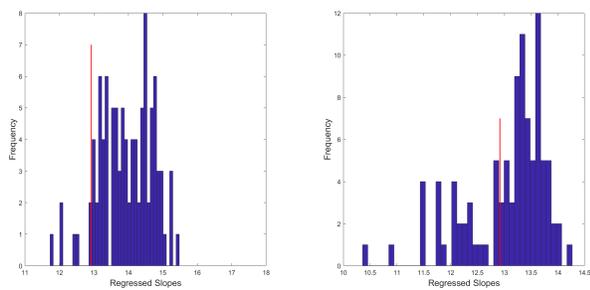


Figure 10: Histogram of regressed slopes taken from 100 random cities with weights in P chosen randomly (left) and proportionally (right). In both cases, $\gamma = 0.01$. The red line denotes the value predicted by the linear spreading speed

If one does have an estimate for the weights of the network, then the heat kernel method can once again be employed to estimate arrival times. We comment briefly on the dynamics for proportional and random weights. Generally speaking, arrival times for proportional P are more

accurate. This makes intuitive sense, as most pairs of nodes have shortest paths connecting them through high degree nodes and proportional weights pushes a higher proportion of travelers along these routes. In contrast, random P matrices can lead to large errors even for small values of γ , depending on the realization of the network. We show the results of one such problematic example in Figure 11. Here we have $\gamma = 0.001$. The initial infection city there is Boriziny, Madagascar, chosen due to the large deviations between predicted and observed arrival times. The reason for the deviations in this case are easy to locate. There are only two connections to Boriziny and the random weights selected in this realization give much larger weight to one of them. Therefore any city whose shortest path to Boriziny connects through the city with less weight will exhibit errors. However, since the two airports connecting to Boriziny are themselves connected by a flight means that the most probable path requires one more flight than the shortest path. Including one more term in the heat kernel approximation leads to a dramatic increase in the accuracy of the approximation, see Figure 11.

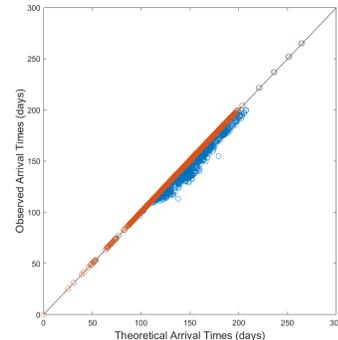


Figure 11: Simulated arrival times compared to one and two-term heat kernel approximations plotted on the line $y = x$ for an infection started in Boriziny, Madagascar with $\gamma = 0.001$. One term approximation is shown in blue and two is shown in orange.

V. CONCLUSIONS

We have presented two estimates for arrival times of an infectious disease in a SIR meta-population model defined on a global airline network. Both estimates are valid in the limit as the rates of diffusion are dominated by the rates of reaction. The linear spreading speed estimate provides a single number for the entire network and can be used to determine the leading order dependence of the arrival times on system parameters. The heat kernel estimate requires more information regarding the network structure and provides a formula, in terms of the Lambert W function, for arrival times between any two nodes in the network.

In the direction of more realistic models of epidemics it would be interesting to understand the role of heterogeneities in system parameters, stochastic effects, time varying connections and finite size effects. For example, city dependent reaction and recovery rates α_n, β_n could likely be incorporated into the heat kernel estimate in a rather straightforward way.

Our methods are not limited to the model considered here, but rely on the linear determinacy of the system. This is a property shared by many systems and the methods here could likely be employed to determine arrival times in other systems and for other networks. We mention the Fisher-KPP equation, which governs the dynamics of a species in a spatially extended environment is linearly determinate; see for example [4, 15]. Of course,

there are many examples of systems which are not linearly determinate and for which the methods developed here would not directly apply. It would be interesting to see whether these methods could be adapted to those classes of problems.

Acknowledgments

This research was conducted as part of a NSF sponsored REU that was held at George Mason University during the summer of 2017. All authors received support from the NSF (DMS-1516155).

-
- [1] A. Barrat, M. Barthélemy, and A. Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, Cambridge, 2008.
- [2] V. Belik, T. Geisel, and D. Brockmann. Natural human mobility patterns and spatial spread of infectious diseases. 1, 03 2011.
- [3] D. Brockmann and D. Helbing. The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342(6164):1337–1342, 2013.
- [4] R. Burioni, S. Chibbaro, D. Vergni, and A. Vulpiani. Reaction spreading on graphs. *Phys. Rev. E*, 86:055101, Nov 2012.
- [5] Y.-Y. Chen, J.-S. Guo, and F. Hamel. Traveling waves for a lattice dynamical system arising in a diffusive endemic model. *Nonlinearity*, 30(6):2334, 2017.
- [6] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani. Predictability and epidemic pathways in global outbreaks of infectious diseases: the sars case study. *BMC Medicine*, 5(1):34, Nov 2007.
- [7] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7):2015–2020, 2006.
- [8] P. Crépey, F. P. Alvarez, and M. Barthélemy. Epidemic variability in complex networks. *Phys. Rev. E*, 73:046131, Apr 2006.
- [9] O. Diekmann and J. A. P. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Wiley, 1 edition, 2000.
- [10] A. Gautreau, A. Barrat, and M. Barthélemy. Arrival time statistics in global disease spread. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(09):L09001, 2007.
- [11] A. Gautreau, A. Barrat, and M. Barthélemy. Global disease spread: Statistics and estimation of arrival times. *Journal of Theoretical Biology*, 251(3):509 – 522, 2008.
- [12] R. F. Grais, J. Hugh Ellis, and G. E. Glass. Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European Journal of Epidemiology*, 18(11):1065–1072, Nov 2003.
- [13] R. Guimerà, S. Mossa, A. Turttschi, and L. A. N. Amaral. From the Cover: The worldwide air transportation network: Anomalous centrality, community structure, and cities’ global roles. *Proceedings of the National Academy of Science*, 102:7794–7799, May 2005.
- [14] J. Hindes, S. Singh, C. R. Myers, and D. J. Schneider. Epidemic fronts in complex networks with metapopulation structure. *Phys. Rev. E*, 88:012809, Jul 2013.
- [15] A. Hoffman and M. Holzer. Invasion fronts on graphs: the Fisher-KPP equation on homogeneous trees and Erdos-Rényi graphs. *preprint*, 2017.
- [16] M. Holzer and A. Scheel. Criteria for pointwise growth and their role in invasion processes. *J. Nonlinear Sci.*, 24(4):661–709, 2014.
- [17] Y. Hosono and B. Ilyas. Traveling waves for a simple diffusive epidemic model. *Math. Models Methods Appl. Sci.*, 5(7):935–966, 1995.
- [18] L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *Proceedings of the National Academy of Sciences of the United States of America*, 101(42):15124–15129, 2004.
- [19] F. Iannelli, A. Koher, D. Brockmann, P. Hövel, and I. M. Sokolov. Effective distances for epidemics spreading on complex networks. *Phys. Rev. E*, 95:012313, Jan 2017.
- [20] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics. ii. the problem of endemicity. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 138(834):55–83, 1932.
- [21] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.
- [22] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Rev. Mod. Phys.*, 87:925–979, Aug 2015.
- [23] M. A. Porter and J. P. Gleeson. *Dynamical systems on networks*, volume 4 of *Frontiers in Applied Dynamical Systems: Reviews and Tutorials*. Springer, Cham, 2016. A tutorial.
- [24] L. A. Rvachev and I. M. Longini. A mathematical model for the global spread of influenza. *Mathematical Biosciences*, 75(1):3 – 22, 1985.
- [25] W. van Saarloos. Front propagation into unstable states. *Physics Reports*, 386(2-6):29 – 222, 2003.
- [26] Z. Wu, L. A. Braunstein, V. Colizza, R. Cohen, S. Havlin, and H. E. Stanley. Optimal paths in complex networks

with correlated weights: The worldwide airport network.

Phys. Rev. E, 74:056104, Nov 2006.