LETTERS IN BIOMATHEMATICS
An International Journal

Intercollegiate Biomathematics Alliance

RESEARCH ARTICLE

OPEN ACCESS

Statistical decomposition of cumulative epidemiological curves into autochthonous and imported cases

Brandon A. Lieberthal[a], Aiman Soliman[b], Allison M. Gardner[a]

[a]School of Biology and Ecology, University of Maine, 5722 Deering Hall, Grove Street Extension, Orono, ME 04469, USA; [b]University of Illinois at Urbana-Champaign, National Center for Supercomputing Applications, 1205 W. Clark Street, Urbana, IL 61801, USA

ABSTRACT

As a pathogen spreads throughout a human population network, understanding the time span between the first reported infection case and the establishment of local transmission relies on the ability to decompose infection incidence into local and travel cases, depending on whether the infected individual was exposed to the pathogen in their location of residence or elsewhere. However, most case data reported to public health agencies do not distinguish between local and travel-associated cases, hampering analysis of the critical early stages of the epidemic spread. We introduce an algorithm, based on the shape of the cumulative incidence curve, to estimate the time a pathogen takes to become locally established, based on the pathogen's transmission and recovery rates and the network connectivity of the human population. This algorithm can predict the onset of an epidemic without considering any future case data, making it useful for tracking epidemics as they occur.

ARTICLE HISTORY
Received April 1, 2020
Accepted August 19, 2020

KEYWORDS
disease outbreak,
human mobility networks,
metapopulation SIR,
local establishment time

1 Introduction

The 21st century presents public health challenges in the form of increasing risk of global epidemics, as an increase in human mobility now is the major driver of infectious disease emergence at domestic and global scales (Cohen, 2000). These include not just contagious diseases such as influenza (Viboud et al., 2006), but also vector-borne diseases such as Zika, chikungunya, and dengue (Lounibos, 2002; Benedict et al., 2007), that are spread primarily via the movement of infected humans. Understanding the influence of human mobility and environmental factors on spread of epidemics relies on the ability to distinguish between autochthonous (i.e. locally acquired) and imported incidence cases as these outbreaks are occurring (Nah et al., 2016). Therefore, development of techniques to differentiate between local and travel associated transmission may offer new insights into the early stages of an epidemic that are critical to public health management decision making (Adams and Kapan, 2009).

Typically, an infectious disease enters a human population center (e.g. a city, state, or country) via the introduction of a small number of infectious individuals. The shape of the incidence curve (i.e., the graph of the cumulative number of cases over time) is initially linear, referred to as the "noisy regime" in Pastor-Satorras et al. (2015). This regime is characterized by a seemingly random assortment of mostly imported cases (i.e., infected individual entering the population) and a few autochthonous, or local cases (i.e., transmission occurring within the population) (Figure 1). After a short period of time, the disease spreads within the local population, the growth of the incidence curve becomes exponential, and the outbreak becomes an epidemic. A variety of analytical (Valdano et al., 2018), numerical (Zhang and Jin, 2011), and empirical (Riou et al., 2017) research has examined the environmental and social conditions that promote the outbreak of epidemics, in terms of the R_0 factor (i.e., the average number of secondary cases arising from a single primary case). Recent research has also explored how temporal and spatial heterogeneity in human population density affects the shape of the outbreak curve (Chowell et al., 2016). However, little research has considered the dynamics of the noisy regime, or the duration between the introduction of the first infected individual and the development of the epidemic, a critical period for effective containment of the outbreak. The structure of the human movement network may alter the dynamics of the noisy regime, in terms of how an outbreak develops from imported cases, and merits thorough study.

CONTACT Brandon A. Lieberthal brandon.lieberthal@maine.edu

Human movement can be modeled with a simple network structure, in which each node represents a population center, and edges represent travel routes between nodes. Nodes are typically weighted by their population size, or subpopulation, and edges are weighted by the number of individuals traveling between nodes per unit time. Previous research demonstrates that the structure of the human mobility network, including heterogeneity and mobility rates, has a significant effect on other metrics of epidemic severity such as R_0, peak prevalence, and epidemic duration. For example, highly clustered networks tend to have longer epidemic durations but lower maximum rates of infection, and networks with high rates of community mixing can have up to double the total number of cases (Carnegie, 2018). In an infectious contact network structure, the assumption that human contact is homogeneous may lead to overly conservative estimates of R_0 (Trapman et al., 2016). Even networks with similar rates of clustering and connectivity may exhibit widely varying epidemic properties, such as epidemic duration and peak prevalence, depending on the parameters used to construct the models (Ball et al., 2013).

To understand the dynamics of the noisy regime, it is imperative to be able to distinguish between local cases and imported cases. This is a major challenge to empirical study of the early dynamics of an outbreak, because in the case of most epidemic outbreaks, for matters of privacy, government officials only report the total number of infected individuals within their jurisdictions (e.g. a public health district or municipality) and do not distinguish between local and imported cases. Alongside other uncertainties such as reporting bias and inconsistency, it is therefore difficult to empirically study the noisy regime, but a better understanding of the underlying dynamics would help to inform the influence of human mobility on the severity of an epidemic within a population.

Here, we test the hypothesis that by analyzing the shape of the cumulative incidence curve, we can infer the underlying dynamics of local and imported cases. Specifically, while the cumulative incidence curve is dominated by imported cases, the shape of the curve is linear or sub-linear ($I(t) \sim t^n, n \leq 1$ or $I(t) \sim \log(t)$), because we expect the rate of imported cases to be relatively constant with time. When local cases begin to dominate, the total incidence curve becomes super-linear, typically a polynomial of degree greater than 1, or exponential in idealized circumstances. There are two advantages for using this method to detect an epidemic outbreak, defined as super-linear growth in total cases. First, it depends only on properties of the cumulative incidence curve, even if the relative proportions of local and imported cases are unknown. Second, the algorithm proposed in this article can detect the establishment of local transmission absent any future incidence data, making it a reliable prediction method for epidemics in progress.

In this paper, we use analytical, numerical, and empirical analyses of epidemics to study the duration in time between the introduction of the first infected individual and the onset of super-linear incidence growth within a population, which we refer to as the "local establishment time." First, we introduce an analytical model of an epidemic on a human population network and show that the local establishment time can be predicted by the disease's transmission and recovery rate, human diffusion, and each node's degree of connectivity. We then simulate these outbreaks on a variety of randomized networks to study the effects of network graph structure on the local establishment time. Finally, we study the western hemisphere chikungunya outbreak of 2014–2015, a recent case study for which published, public human case data are decomposed into local and imported cases, to validate our hypotheses.

1.1 Introduction to SIR Models

The basic SIR (Susceptible-Infected-Recovered) model is a mathematical framework used in epidemiology to describe the dynamics of an infectious disease. Individuals from an at-risk population of size N are classified among three states (i.e., susceptible, S; infected, I; or recovered, R). Differential equations describe the rate of transition between states based upon transmission and recovery rates derived from empirical data. A metapopulation SIR model is an elaboration of the basic model that considers a network of population centers or nodes, each with its own set of SIR equations, and the rates of human migration between nodes (Anderson and May, 1984).

Consider an undirected mathematical graph of nodes labeled $i = 1, \ldots, n$. The network as a whole represents the metapopulation network, whereas each node represents a local human population (Ray et al., 1997). Parameters for the simulation include an initial condition, an infection rate β (the rate at which the disease travels between individuals), a recovery rate μ (the rate at which an individual recovers from the disease), and a human mobility matrix M_{ij} (the rate of movement from node i to node j).

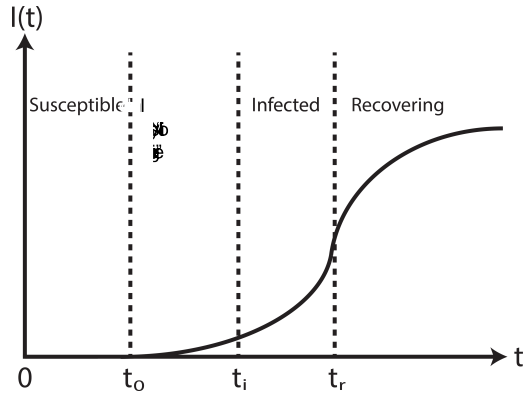

Figure 1: An illustration of the cumulative epidemic curve for a population, and the classification of the population into one of four different states based on the curve.

Thus, the metapopulation SIR model for a network of size n can be expressed as a set of $3n$ ordinary differential equations as:

$$\begin{aligned}
\frac{\mathrm{d}S_i}{\mathrm{d}t} &= -\beta\frac{S_iI_i}{N_i} + \sum_{j=1}^{n} M_{ji}S_j - \sum_{j=1}^{n} M_{ij}S_i \\
\frac{\mathrm{d}I_i}{\mathrm{d}t} &= \beta\frac{S_iI_i}{N_i} - \mu I_i + \sum_{j=1}^{n} M_{ji}I_j - \sum_{j=1}^{n} M_{ij}I_i \\
\frac{\mathrm{d}R_i}{\mathrm{d}t} &= \underbrace{\mu I_i}_{\text{local transmission}} + \underbrace{\sum_{j=1}^{n} M_{ji}R_j - \sum_{j=1}^{n} M_{ij}R_i}_{\text{host mobility}}
\end{aligned} \tag{1}$$

While much of prior research on metapopulation SIR models has focused on the implications of network topology for the epidemic basic reproduction number, $R_0 = \frac{\beta}{\mu}$ (Apolloni et al., 2014), our study investigates how network properties impact the time between the introduction of an infected individual into a network, the arrival of the first infected individual to a node, and the establishment of local transmission within a node. Specifically, we apply a more holistic approach in which rather than classifying individuals as susceptible, infected, or recovered, we classify entire nodes into one of four states based on the cumulative growth of infected individuals (Figure 1).

- At $t = 0$, the first individual within the network graph becomes infected. Any nodes with no infected individuals are labelled as "Susceptible."
- At $t = t_0$, an infected individual enters a node. The epidemic curve is roughly linear, since most cases arise from traveling individuals, and the node is classified as "Introduced." Pastor-Satorras et al. (2015) refers to this period of time as the "noisy regime."
- At $t = t_i$, local transmission of the disease overtakes the dynamics of the epidemic curve within the node. The epidemic curve is now exponential, and the node is classified as "Infected."
- At $t = t_r$, the number of infected individuals approaches a maximum, and the cumulative infection time series becomes convergent. The node is finally classified as "Recovering." The Recovering state is not discussed in this study, because as described in Section 3.3, it is difficult to identify in empirical situations.

Here, we are primarily interested in the time duration between $t = t_0$ and $t = t_i$, the time between the introduction of the epidemic and the onset of local transmission in a specific node, which in this paper we define as the local establishment time. We address three questions: (1) How effectively can we detect the local establishment time based on the properties of the cumulative incidence curve? (2) How do node properties (e.g., node connectivity and centrality) alter the local establishment time of an infectious disease in simulated networks? (3) Do the same principles apply to the empirical example of the spread of chikungunya virus in the western hemisphere? Our findings may be consequential for timely and effective allocation of public health resources during the initial stages of an infectious disease outbreak.

2 Methods

2.1 Metapopulation SIR Model

We use Equation (1) as the foundation for our metapopulation SIR model. In Section 3.2, we experiment with several types of network graphs to test the hypothesis that the distribution of high and low connectivity nodes can affect the local establishment time throughout the network. These networks are described in more detail in that section.

Along with the mathematical graph (or node-edge arrangement), a metapopulation network is also characterized by its mobility matrix, for which M_{ij} represents human movement per unit time from node i to node j. In a traffic dependent mobility rate model, the probability that an individual migrates per unit time is given by the diffusion constant p, with a value between 0 and 1 (Colizza and Vespignani, 2008). The simplest possible mobility matrix is defined as $M_{ij} = \frac{p}{k_i}$, where k_i is the degree of connectivity of node i, but this model unrealistically assumes that all human movement is homogeneous. A more realistic traffic dependent model assumes that migration rates between nodes is heterogeneous, with a mobility matrix $M_{ij} = p\frac{w_0(k_ik_j)^\theta}{Ak_i^{1+\theta}}$, where w_0 and θ depend on the specific mobility system ($\theta = 0.5$ is a common metric for air routes (Barrat et al., 2003)), and A is a calibration factor chosen such that the total fraction of individuals migrating from a node per unit time adds up to p. The advantage of the traffic-dependent model, as discussed in Section 3.1, is its relative ease at deriving analytical solutions to the metapopulation SIR equations. The disadvantage of this model is that for nodal subpopulations to be convergent in time, it is required that the initial population of each node be directly proportional to $k^{1+\theta}$.

An alternative to the traffic dependent mobility matrix is the population dependent mobility matrix, which is designed such that the total population in each node is convergent regardless of the initial conditions. In this model, a fixed number of people migrate between any given pair of nodes per unit time, and the mobility rate is inversely proportional to the population of each node, $M_{ij} = \frac{w_0(k_ik_j)^\theta}{N_i}$. The disadvantage of this model is that finding an analytical solution in the early outbreak limit is difficult.

2.2 Local Establishment Time

The metapopulation SIR model is run in MATLAB, based on a set of $3n$ ordinary differential equations (Keeling and Rohani, 2007). The simulation is set to run for a sufficient duration that the epidemic can spread to every node in the network. As the simulation runs, for each node, the values of S, I, and R are computed as time series. We hypothesize that the cumulative infected cases for each node initially will be linear, as infected individuals enter the node, and then will transition to an exponential curve as the disease becomes established locally. For each node, we set t_0 as the time at which the first infection is reported. We use the deceleration of growth parameter, as defined in Chowell et al. (2017), to describe the shape of the curve with a single numeric value. We can use this parameter to compute the first moment at which the incidence curve becomes super-linear. This point in time is estimated as the moment at which local transmission dominates the dynamics of the incidence curve.

The deceleration of growth parameter, which here we call g, is defined implicitly by the equation

$$\frac{\mathrm{d}I(t)}{\mathrm{d}t} = rI(t)^g, \tag{2}$$

where $I(t)$ is the cumulative number of infection cases, and r and g are fitting parameters. It is straightforward to demonstrate that $I(t)$ is sub-linear if $g < 0$, linear if $g = 0$, super-linear (e.g. quadratic, cubic, etc.) if $g > 0$, and exponential if $g = 1$.

The deceleration of growth parameter at time t can be explicitly calculated from a data set with the equation

$$g = 1 - \left[\frac{I'(t)}{I(t)} * (t - t_0)\right]^{-1}, \tag{3}$$

where t_0 is the moment in which the first infection case is reported in the subpopulation. This equation is convenient as, unlike in the original definition, it requires no curve fitting algorithm.

2.3 Empirical Data

The chikungunya virus, following a large outbreak in Kenya in 2004, underwent a rapid global expansion across Africa and Asia (Staples et al., 2009), with localized outbreaks in Europe (Tomasello and Schlagenhauf, 2013), extensive outbreaks throughout the Caribbean, Central and South America, and Mexico, and locally-acquired cases in Florida the following year (Weaver and Forrester, 2015). As *Aedes* mosquitoes only have a flight range of at most a few kilometers (Briegel et al., 2001), the spread of chikungunya and other epidemics throughout the world, especially at large spatial scales, are driven primarily by human movement (Moore et al., 2018). An important question in disease ecology concerns the drivers of disease spread and identifying risk factors, particularly in the early stages of an outbreak, to inform allocation of limited public health resources such as vector control, education efforts, and vaccinations (Steele et al., 2016).

The 2014 chikungunya outbreak throughout the western hemisphere was chosen for empirical analysis because the data set is one of the few publicly available data sets that explicitly distinguishes between local and imported cases (PAHO, 2015). Infection cases were reported to the Pan-American Health Organization by 50 nations or territories throughout the western hemisphere on a weekly basis. Cases are classified as "Confirmed" if the patient was positive for chikungunya virus as determined by laboratory testing, "Suspected" if the patient exhibited the symptoms of chikungunya virus but was not tested, and "Imported" if the patient recently had traveled in another affected region. For the purposes of this analysis, the number of cases per geographic unit per time step is calculated as the sum of confirmed and suspected cases.

Because some countries stopped reporting cases after the outbreak of an epidemic within their jurisdiction, these data are not practical for evaluating the duration or peak prevalence of the epidemic (Lipsitch et al., 2015). In addition, some countries reported cases at irregular intervals, which may give the false impression that the epidemic halted and restarted over the course of the year. However, these data are still useful for measuring the local establishment time within each region, because reporting tends to be high during the early stages of a disease outbreak.

The data are publicly available as weekly reports, and from each report the number of cases, divided into Suspected, Confirmed, and Imported, were manually transcribed into a spreadsheet. No data cleaning was necessary, besides fixing obvious typographical errors such as transposed digits, and countries with ten or fewer confirmed cases in 2014 were not considered in our analysis because there were not enough data points to precisely compute the shape of the curve.

International tourism data for each country in 2014 were obtained by the World Bank (2019). This data set refers to the total number of arrivals of non-resident visitors at national borders, including nationals residing abroad, per country per year.

3 Results

3.1 Analytical Model

The analytical model described in this section expands upon the analysis conducted by Colizza and Vespignani (2008), which solves the metapopulation SIR equations under the assumption that each node's population and traffic is dependent on its degree of connectivity, and in the asymptotic limit of an emerging outbreak. Statistically, a node of degree k is expected to have an initial epidemic outbreak incidence curve function of

$$I_k(t) = B\frac{k^{1+\theta}}{\langle k^{1+\theta}\rangle}e^{(\beta-\mu)t} + C_k e^{[(1-p)(\beta-\mu)-p]t} \tag{4}$$

where k is the node's degree of connectivity, θ is the heterogeneity of movement constant, β and μ are the infection and recovery rates, p is the diffusion constant (assumed to be the same for all nodes), and B and C_k are parameters determined by the initial conditions. If it is assumed that the initial infected population is equally distributed across the entire grid, then $B = \bar{I}(0)$ and $C_k = \bar{I}(0)(1 - k^{1+\theta}/\langle k^{1+\theta}\rangle)$, where $\bar{I}(0)$ represents the average number of infected individuals among all infected nodes at time $t = 0$.

To develop an analytical model to estimate local establishment time, our goal is to study the logistic progression of an epidemic in an individual node, on the assumption that each of its adjacent nodes follow the statistical incidence curve expressed in Equation (4). Our strategy is to consider an individual node of degree k that is initially uninfected. This node has neighbors $j = 1, 2, \ldots, k$ each with degree k_j. We assume that the neighboring nodes' infection rates can be estimated by Equation (4). We define $t = 0$ as the moment when these neighboring nodes were first infected, and $t = t_0$ as the moment when the individual node in question becomes first infected. Therefore, the infection rate of this individual node, indicated by J, is given by

$$\frac{\mathrm{d}J}{\mathrm{d}t} = (\beta - \mu)J(t) + \sum_j (M_{ji}I_j - M_{ij}J) \tag{5}$$

Substituting in I_{k_j} from Equation (4) for $I_j(t)$ in Equation (5), this equation becomes

$$\frac{\mathrm{d}J}{\mathrm{d}t} = \left(\beta - \mu - \sum_j M_{kk_j}\right)J(t) + \sum_j M_{k_jk}\left[B\frac{k^{1+\theta}}{\langle k^{1+\theta}\rangle}e^{(\beta-\mu)t} + C_k e^{[(1-p)(\beta-\mu)-p]t}\right] \tag{6}$$

Where M_{kk_j} is the traffic dependent model's prediction of mobility rates between a node of degree k and a node of degree k_j. Note that with the assumption of constant diffusion, we can make some simplifications to the terms that are dependent on k but not t:

$$\sum_j M_{kk_j} = p \tag{7}$$

$$\sum_j M_{k_j k} B \frac{k_j^{1+\theta}}{\left\langle k_j^{1+\theta} \right\rangle} = \bar{I}(0) * p \tag{8}$$

$$\sum_j M_{k_j k} C_k = \sum_j M_{k_j k} \bar{I}(0) \left(1 - \frac{k_j^{1+\theta}}{\left\langle k_j^{1+\theta} \right\rangle} \right) = \bar{I}(0) \left(\sum_j M_{k_j k} - p \right) \tag{9}$$

In the traffic dependent model, the $\sum_j M_{k_j k}$ term evaluates to

$$\sum_j M_{k_j k} = \sum_j p \frac{w_0 (k_j k)^\theta}{A k_j^{1+\theta}} = p \frac{k^\theta}{\langle k \rangle^\theta} \sum_j \frac{1}{k_j} \tag{10}$$

Then Equation (6) simplifies to

$$\frac{\mathrm{d}J}{\mathrm{d}t} = (\beta - \mu - p) J(t) + \bar{I}(0) p e^{(\beta-\mu)t} + \bar{I}(0) p \left(\frac{k^\theta}{\langle k \rangle^\theta} \sum_j \frac{1}{k_j} - 1 \right) e^{[(1-p)(\beta-\mu)-p]t} \tag{11}$$

The solution to this equation is

$$J(t) = K e^{(\beta-\mu-p)(t-t_0)} + \bar{I}(0) e^{(\beta-\mu)t} + \frac{\bar{I}(0)}{\beta - \mu} \left(1 - \frac{k^\theta}{\langle k \rangle^\theta} \sum_j \frac{1}{k_j} \right) e^{[(1-p)(\beta-\mu)-p]t} \tag{12}$$

where K is a constant chosen such that $J(t_0) = 0$. Specifically,

$$K = -\bar{I}(0) e^{(\beta-\mu)t_0} - \frac{\bar{I}(0)}{\beta - \mu} \left(1 - \frac{k^\theta}{\langle k \rangle^\theta} \sum_j \frac{1}{k_j} \right) e^{[(1-p)(\beta-\mu)-p]t_0} \tag{13}$$

This equation simplifies to

$$\begin{aligned} J(t) = \bar{I}(0) \left[e^{(\beta-\mu)t} - e^{(\beta-\mu)t_0} e^{(\beta-\mu-p)(t-t_0)} \right] \\ + \frac{\bar{I}(0)}{\beta - \mu} \left(\frac{k^\theta}{\langle k \rangle^\theta} \sum_j \frac{1}{k_j} - 1 \right) \left[e^{[(1-p)(\beta-\mu)-p]t_0} e^{(\beta-\mu-p)(t-t_0)} - e^{[(1-p)(\beta-\mu)-p]t} \right] \end{aligned} \tag{14}$$

This solution is a sum of six exponential curves, each proportional to the constant $\bar{I}(0)$. Very roughly speaking, the first set of terms represents local infection cases, while the second set of terms represent travel cases, either imported or exported. The exact values of β, μ, and p depend on the choice of time unit, however, β and μ are typically an order of magnitude smaller than p (Cross et al., 2005). Therefore, the first term in Equation (14), $\bar{I}(0)e^{(\beta-\mu)t}$ is the only term with a positive exponent, and will dominate the dynamics of the early outbreak fairly quickly. The question is pinpointing the moment in time at which this occurs.

The sign of the second term depends on the sign of $\left(\frac{k^\theta}{\langle k \rangle^\theta} \sum_j \frac{1}{k_j} - 1 \right)$, which we call the "exchange term." The exchange term tends to be positively correlated with k and negatively correlated with the clustering coefficient. In all of the randomly generated scale-free network models that we tested, this term will be greater than zero if k is greater than the mean k, approximately zero if k is equal to the mean, and less than zero if k is less than the mean. This is appropriate, because if we briefly assume that $k_j = \langle k \rangle$, then this term simplifies to $\left(\frac{k}{\langle k \rangle} \right)^{1+\theta} - 1$. This implies that in high connectivity nodes, the net travel cases during the initial outbreak will be positive, and in the low connectivity nodes, they will be negative. Although the travel case term will always asymptotically approach zero over time, the sign of this term has implications for the short-term dynamics of the epidemic curve. An example of this is shown in Figure 2. Note the slightly different concavity of the initial epidemic curve depending on whether the node's degree of connectivity is greater than or less than the mean across the network.

In order to predict the local establishment time, our goal is to determine the moment in time, designated as t_i, when the epidemic curve described by Equation (14) transitions from a linear to a super-linear pattern. Since this is at heart an asymptotics problem, there is not an explicit solution to this problem. However, the shape of the epidemic curve can be characterized by a

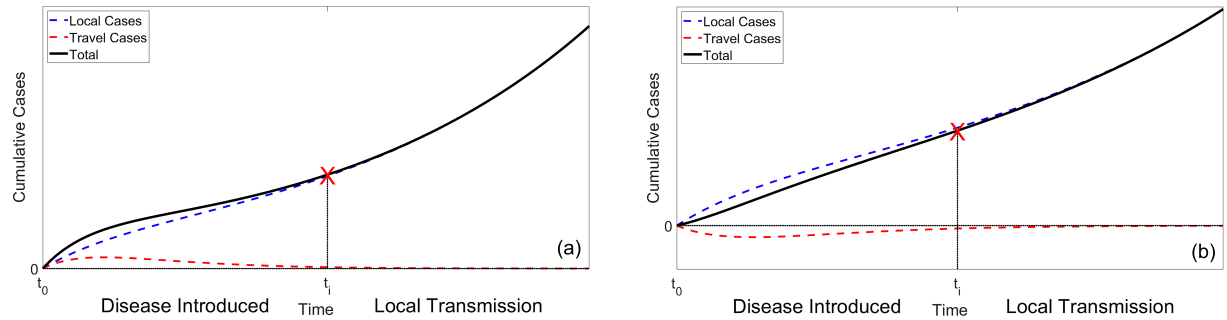

Figure 2: An illustration of the epidemic outbreak curve, decomposed into local and travel cases, on a node with degree of connectivity (a) greater than the mean and (b) less than the mean. The local establishment time is labeled as the first moment when the total outbreak curve becomes super-linear.

deceleration of growth parameter, defined in Equation (2), and this transition occurs when the deceleration of growth parameter first becomes positive.

Tracking the evolution of the epidemic curve shape over time is therefore as simple as computing the deceleration of growth parameter over time. It is also straightforward to demonstrate that with common metapopulation SIR parameters (for example, $\beta = 0.04$, $\mu = 0.02$, $p = 0.5$), that the deceleration of growth parameter is initially negative as the curve is sub-linear. After a short period of time, when the local infection case term begins to dominate the dynamics of the curve, the deceleration of growth parameter increases past 0, at which point the curve transitions from sub-linear to linear to super-linear. Eventually, the deceleration of growth parameter approaches a value close to 1, at which point the epidemic curve is exponential or nearly exponential.

With this in mind, **we define the local establishment time, t_i, as the moment when the deceleration of growth parameter becomes greater than zero.**

Fortunately, there are properties of the deceleration of growth parameter that make this easy to solve. We take the derivative of Equation (2) and make a substitution to find

$$I''(t) = rgI(t)^{g-1}I'(t) = r^2 gI(t)^{2g-1} \tag{15}$$

Note that since r^2 and $I(t)$ are always positive, the second derivative of $I(t)$ has the same sign as g, the deceleration of growth parameter. Therefore, the local establishment time can also be defined as the moment when the second derivative of the cumulative incidence curve transitions from negative to positive.

In the limiting case of $k = k_j = \langle k \rangle$, the second term of Equation (14) vanishes, and the equation is reduced to

$$J(t) = \bar{I}(0)\left[e^{(\beta-\mu)t} - e^{(\beta-\mu)t_0}e^{(\beta-\mu-p)(t-t_0)}\right] \tag{16}$$

Setting the second derivative of $J(t)$ equal to 0:

$$J''(t) = \bar{I}(t)\left[(\beta-\mu)^2 e^{(\beta-\mu)t} - (\beta-\mu-p)^2 e^{(\beta-\mu)t_0}e^{(\beta-\mu-p)(t-t_0)}\right] = 0 \tag{17}$$

we find an exact solution for the local establishment time:

$$t_i - t_0 = \frac{2}{p}\log\left(\frac{p}{\beta-\mu} - 1\right) \tag{18}$$

This value is a baseline estimate for the local establishment time. We expect the local establishment time to be greater for low connectivity nodes and smaller for high connectivity nodes. In the case of $\beta = 0.04$, $\mu = 0.02$, $\theta = 0.5$, and $\langle k \rangle = k$, we numerically computed $t_i - t_0$ for a range of values of k and p. Our results are shown in Figure 3. As expected, the local establishment time is negatively correlated with k and with p, although both of these relationships are quickly convergent.

3.2 Numerical Simulation

We use a numerical simulation of a metapopulation SIR model to demonstrate that the deceleration of growth parameter can be used to analyze the *total* (local and travel) cases to predict when the local cases are dominating the epidemic curve. For these

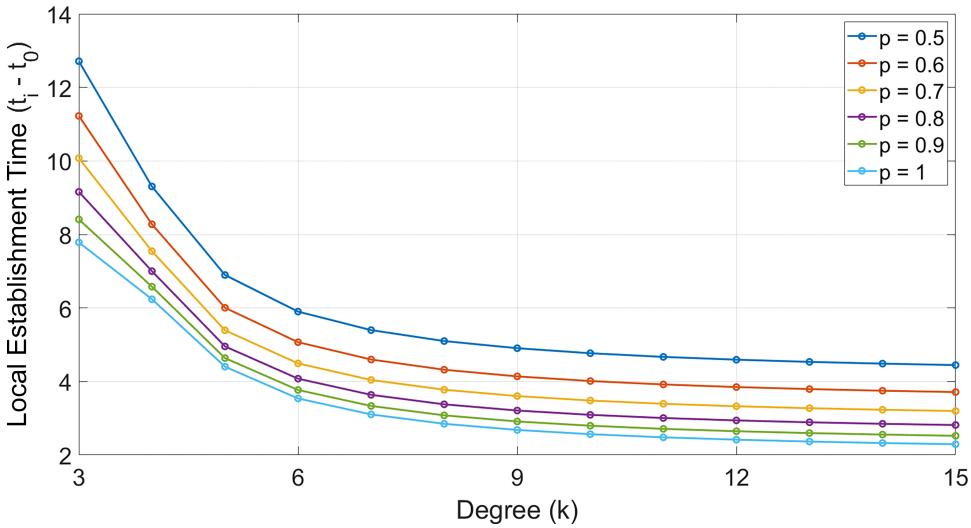

Figure 3: Numerical solution to the local establishment time for a range of values of k and p, with $\beta = 0.04$, $\mu = 0.02$, $\theta = 0.5$, and $\langle k \rangle = k$.

simulations, we can verify our predictions with the decomposed local and travel epidemic curves, but this method will also prove useful when these data are not available.

Our test cases include a scale-free network, with 5000 nodes and degree distribution law $P(k) \sim k^{-3}$ (Albert and Barabási, 2002), a small-world model with average connectivity of 9 (Watts and Strogatz, 1998), and a Delaunay triangulation model consisting of nodes randomly arranged around a center point with an exponential distance distribution (Lee and Schachter, 1980). To briefly summarize these networks, the scale-free model is characterized by a few highly connected hub nodes, the small-world model has random edge connections across the network to increase centrality, and the triangulation model has all adjacent nodes connected to increase clustering. Each of these simulations was run with $\beta = 0.04$, $\mu = 0.02$, $p = 0.5$, $\theta = 0.5$, with a traffic-dependent mobility model with $\tilde{N} = 1000$. For each network, nodes are categorized according to their degree of connectivity k and the mean degree of connectivity for that node and all of its neighbors, $\langle k \rangle$.

After the simulation is run, the evolving deceleration of growth parameter is computed for the total incidence curve of each node. The local establishment time, the time between the first infected case and the moment at which the deceleration of growth parameter exceeds zero, is recorded. To reiterate, this metric has two practical advantages: first, it only relies on the total incidence curve, even if the decomposed local and imported curves are not known, and second, it can be computed in real time without knowing any future data.

For these metapopulation SIR simulations, the local and imported incidence curves, designated $I_L(t)$ and $I_I(t)$, are computed by decomposing the differential equation for $I_i(t)$ in Equation (1) as:

$$\begin{aligned} \frac{\mathrm{d}I_i}{\mathrm{d}t} &= \frac{\mathrm{d}I_{L,i}}{\mathrm{d}t} + \frac{\mathrm{d}I_{I,i}}{\mathrm{d}t} \\ \frac{\mathrm{d}I_{L,i}}{\mathrm{d}t} &= \beta \frac{S_i I_i}{N_i} \\ \frac{\mathrm{d}I_{I,i}}{\mathrm{d}t} &= \sum_{j=1}^{n} M_{ji} I_j - \sum_{j=1}^{n} M_{ij} I_i \end{aligned} \tag{19}$$

For each node, the **local crossover time** is defined as the moment at which the local incidence curve first exceeds the imported incidence curve. The local establishment time and the local crossover time are compared based on each node's values of k and $\langle k \rangle$. We compare the local establishment time and the local crossover time to compare characterizations of the disease outbreak based on the shape of the incidence curve or the number of infected individuals.

3.2.1 Results

The results of each simulation are shown in Figure 4. For each network model, the left hand side shows the total, local, and imported incidence curves of 9 randomly selected nodes, sorted by k. Note that the time axis is defined relative to the moment

of first introduction to each node. The right hand side shows a scatter plot of the local establishment time and the local crossover time for each node, colored by the value of $k^\theta/\langle k\rangle^\theta$.

By and large the analytical model presented in Section 3.1 proved to be a reliable prediction of the dynamics of the incidence curve in each node. However, the analytical model overlooks the intricacies of mobility rates between nodes, especially nodes whose value of k differs considerably from its neighbors.

The scale-free network has the least equitable distribution of nodal degree, and therefore has the most variance in local establishment time among nodes (Figure 4a). There is a roughly inverse linear relationship between the natural log of the degree and the local establishment time. Observing individual nodes, we observe three main dynamics regarding the local and imported incidence curves.

1. The local and imported incidence curves both grow linearly, but the local incidence curve exceeds the imported incidence curve almost immediately. This occurs in nodes with small k and small $\langle k\rangle$. In this simulation the local establishment time tends to be close to 40, and the local crossover time is nearly 0.

2. The imported incidence curve is initially super-linear but decays quickly. This occurs in nodes with low k and high $\langle k\rangle$. Essentially, a large number of infected individuals move into this node, faster than local growth. The total incidence curve is super-linear to start, becomes sub-linear briefly, then becomes super-linear again. In these situations, the local establishment time is much longer than the local crossover time.

3. The local and imported incidence curves both grow linearly, but the imported incidence curve has a higher slope. When the local incidence curve becomes super-linear, it eventually overtakes the imported incidence curve. This occurs in nodes with higher k than $\langle k\rangle$, and these types of nodes are most frequent in scale-free models. In these cases, the local establishment time is almost always shorter than the local crossover time.

For the scale-free model, the median local establishment time is 40 and the median local crossover time is 20. Nodes with $k > \langle k\rangle$ tend to have the shortest local establishment time or the longest local crossover time, and those nodes with the highest $k^\theta/\langle k\rangle^\theta$ ratio have a longer local crossover time than local establishment time.

As the small-world model has a more equitable distribution of nodal degree than the scale-free model, the local establishment time was extremely consistent, with every single node having a value between 51 and 54. The local crossover time ranged between 0 and 36, and every node had a shorter crossover time than establishment time. As with the scale-free model, the nodes with the highest $k^\theta/\langle k\rangle^\theta$ ratio also have the longest crossover time.

The Delaunay triangulation model, like the small-world model, has an equitable distribution of nodal degree, but most nodes have a degree of connectivity between 4 and 8. As with the small-world model, most nodes have about the same local establishment time, around 28, with the exception being nodes with degree greater than 8 or nodes on the outskirts of the model. The nodes with high degree tend to have short local establishment times and long local crossover times, and the nodes on the outer boundary of the network tend to have unusually long local establishment times.

3.3 Empirical Analysis

Finally, we used the 2014 chikungunya outbreak throughout the Western hemisphere as an example of the type of analysis that can be performed with the techniques described above. Figure 5a shows the cumulative number of local (confirmed + suspected) and imported chikungunya cases throughout the western hemisphere, on a national spatial scale and a weekly temporal scale. Most countries exhibit the pattern observed in numerical simulations: incidence curves are initially driven by a small number of imported cases, and the epidemic curve is linear. After a short period of time, the growth in local cases overtakes imported cases, and the total epidemic curve has a linear pattern. One major exception is the United States, which has high rates of international entry and poor environmental suitability for the mosquito species that transmit chikungunya, but there are a few other countries, such as Panama and Paraguay, where imported cases are consistently higher than local cases. A few small island countries with irregular reporting patterns are not shown in this figure.

The red 'X' on each plot shows the moment in time that the epidemic curve first becomes super-linear, calculated based on the *total* number of cases, local and imported. This moment tends to be close to the moment when local cases overtakes imported cases. Figure 5b shows the same data as in Figure 5a, but on a log scale, to better illustrate this point. This empirical analysis validates the hypothesis that the deceleration of growth parameter can be used to estimate this local crossover time, even if the available data set does not distinguish between local and imported cases. For the countries that did not reliably report imported cases, such as Belize, Haiti, and Martinique, we can still estimate the crossover time based on this analysis.

It can be observed in Figure 5b that some countries actually had multiple, separate epidemic events, in which the number of local cases jumped sharply. Plotting the deceleration of growth parameter over time, as shown in Figure 5c, is an effective way to record these events. For example, this figure shows that the United States had one prolonged epidemic event, which began

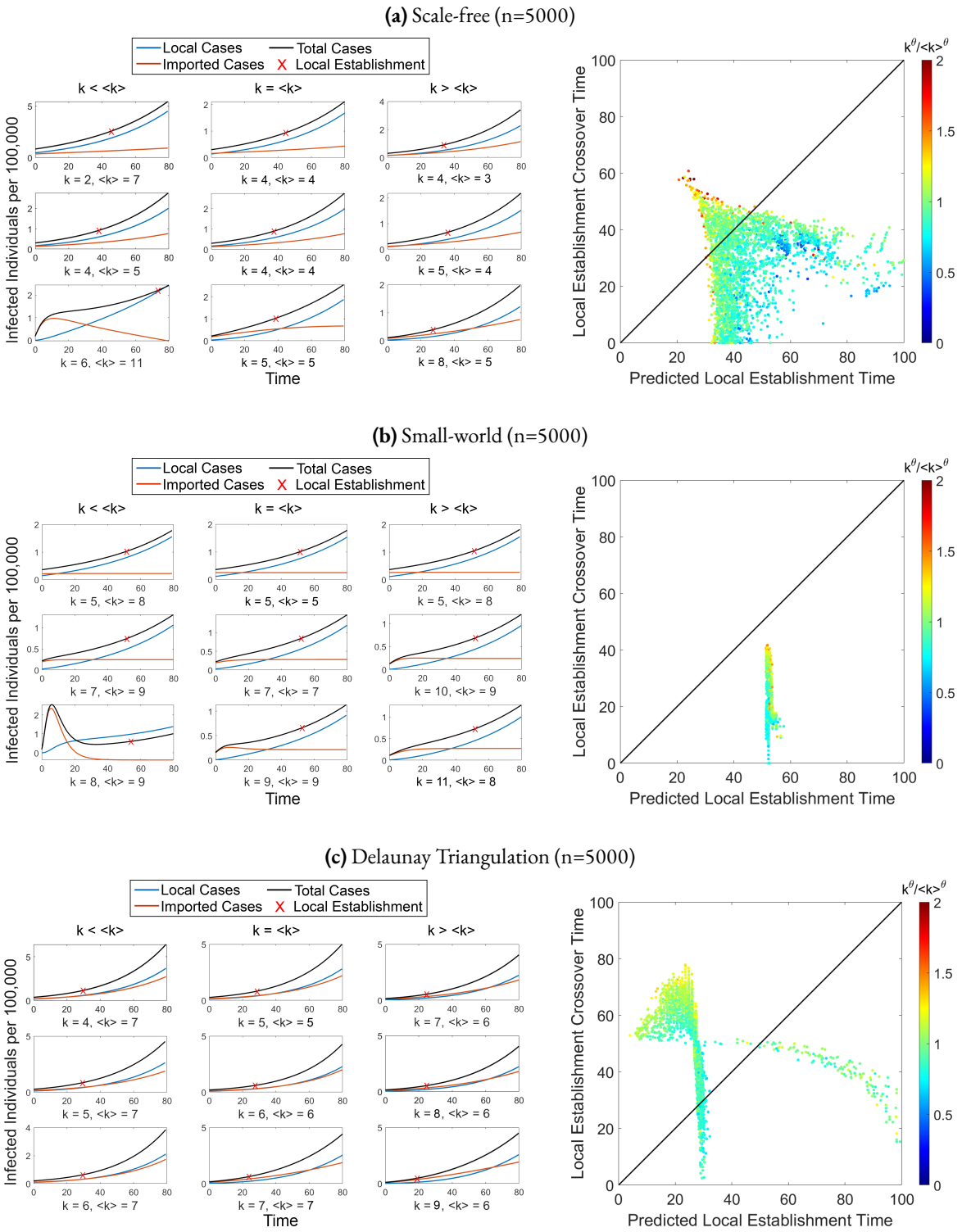

Figure 4: Metapopulation SIR simulation results for a scale-free, small-world, and Delaunay triangulation model with 5000 nodes. For each model, the left figure shows the cumulative incidence curve of the epidemic outbreak in randomly selected nodes. The incidence curve is decomposed into local and imported cases. The red 'X' signifies the local establishment time, the moment in time at which the total incidence curve becomes super-linear. The right figure compares for all 5000 nodes the local establishment time and the local crossover time, the moment in time at which the local incidence curve overtakes the imported incidence curve.

in Week 24 and ended in Week 50. However, Aruba had two clearly identified epidemic events, each lasting 12 and 10 weeks, respectively. Some countries, such as Panama, had three or more such epidemic events, but there is a possibility that this is just an artifact of reporting bias.

The circumstances that results in multiple epidemic events are difficult to determine and dependent on the specific details of each country. However, there is a clear link between international connectivity and the local establishment time within each country. As a proxy for connectivity, we use the each country's total number of international visitors throughout 2014 divided by their resident population. Figure 6 shows an inverse relationship between each country's connectivity and its local establishment time. Countries that have more international visitors than their own populations tends to have short establishment times, typically between 1 to 4 weeks, likely because the disease is introduced through multiple individuals at the same time. There is much more variance among countries with lower rates of tourism, with local establishment times ranging from 1 to 15 weeks, with an average of 8.7 weeks.

4 Discussion

We discuss several types of analysis into an overlooked but critically important aspect of an epidemic outbreak: the span of time between which the first infected individual enters a metapopulation and when local infection cases begin to dominate the dynamics of the epidemic curve. If we assume for now that the environmental and socioeconomic properties are consistent across the population, the dynamics of this initial infection phase are heavily dependent on the structure of the network. As a consequence, certain nodes can be identified as having a high risk factor for epidemics based on their network connectivity relative to their neighbors.

The analytical model presented in Section 3.1, although simplistic, is useful for demonstrating a few important properties of the initial infection phase within a subpopulation. First, the noisy regime is roughly a superposition of six exponential curves, of which two represent local infection cases, two represent the importation of infected individuals from neighboring nodes, and two represent the exportation of infected individuals into neighboring nodes. The only term to have a positive exponent is the local infection $\bar{I}(0)e^{(\beta-\mu)t}$ term, and therefore given sufficient time, this term will eventually dominate the epidemic curve as long as the assumption $\Gamma = \frac{SI}{N} \sim I$ holds. However, the travel infection terms have the coefficient $\frac{\bar{I}(0)}{\beta-\mu}$, which for typical values of β and μ will be orders of magnitude larger than $\bar{I}(0)$. Therefore, for a short period of time we would expect the travel terms to dominate the curve.

The value of the $\left(\frac{k^{\theta}}{\langle k\rangle^{\theta}}\sum_j \frac{1}{k_j} - 1\right)$ term, which we call the exchange term and is constant in time for any specific node, is the main factor in determining the precise dynamics of local and travel infection cases in the very early outbreak. In the idealized limit that all nodes in the network have the same degree of connectivity (for example, a complete graph), the exchange term vanishes, there is always an equal exchange of infected individuals between adjacent nodes, and only the local transmission terms matter. If the exchange term for a particular node happens to be an order of magnitude smaller than $\beta - \mu$, then the result is effectively the same, and the estimated local establishment time from Equation (18) holds.

Outside of this limit, the structure of the network affects the shape of the epidemic curve beyond what the analytical model can predict. Examples of this are shown in Figure 4, which demonstrates how the degree distribution of the network can dramatically affect the distribution of local establishment times. In general, networks with more equitable degree distributions, such as small-world and triangulation models, tend to also have more equitable distributions in local establishment time. It is also true that nodes with higher degree of connectivity tend to have shorter local establishment times but longer local crossover times. In other words, the total epidemic curve becomes super-linear faster than the local infection cases overtakes the imported cases. The opposite tends to be true for nodes with lower than average connectivity, in that the local establishment time is longer than the local crossover time, but this time difference is most stark in the scale-free model. In general, the local establishment time tends to be more homogeneous across the network than the local crossover time.

A fraction of nodes in the small-world model have an unusual dynamic in which the epidemic curve is triggered by a large "bump" in travel cases. Nodes with this quality tend to have a negative exchange term, but there is no apparent correlation between the exchange term and the probability of having this bump. They occur in nodes that are adjacent to several nodes in advanced states of epidemic, and they are triggered by a sudden influx of a large number of infected individuals. These bumps cause the local establishment time to be about 10% longer, as the imported incidence curve has an outsized effect on the dynamics of the epidemic within that node.

The numerical simulations discussed in Section 3.2 demonstrate the predictive and diagnostic utility of the local establishment time. A major advantage of the local establishment time as a metric of infection speed, besides the unnecessity of decomposed local and imported case data, is its remarkable stability across a network. In the scale-free network shown in Figure 4, which follows a $P(k) \sim k^{-3}$ degree distribution, the standard deviation in local establishment time is about 12.2 units, whereas the standard deviation in local crossover time is about 31.0 units. For individual nodes, the local establishment time may underpredict an epidemic for nodes with connectivity less than half the mean of their neighbors, which would typically pertain to

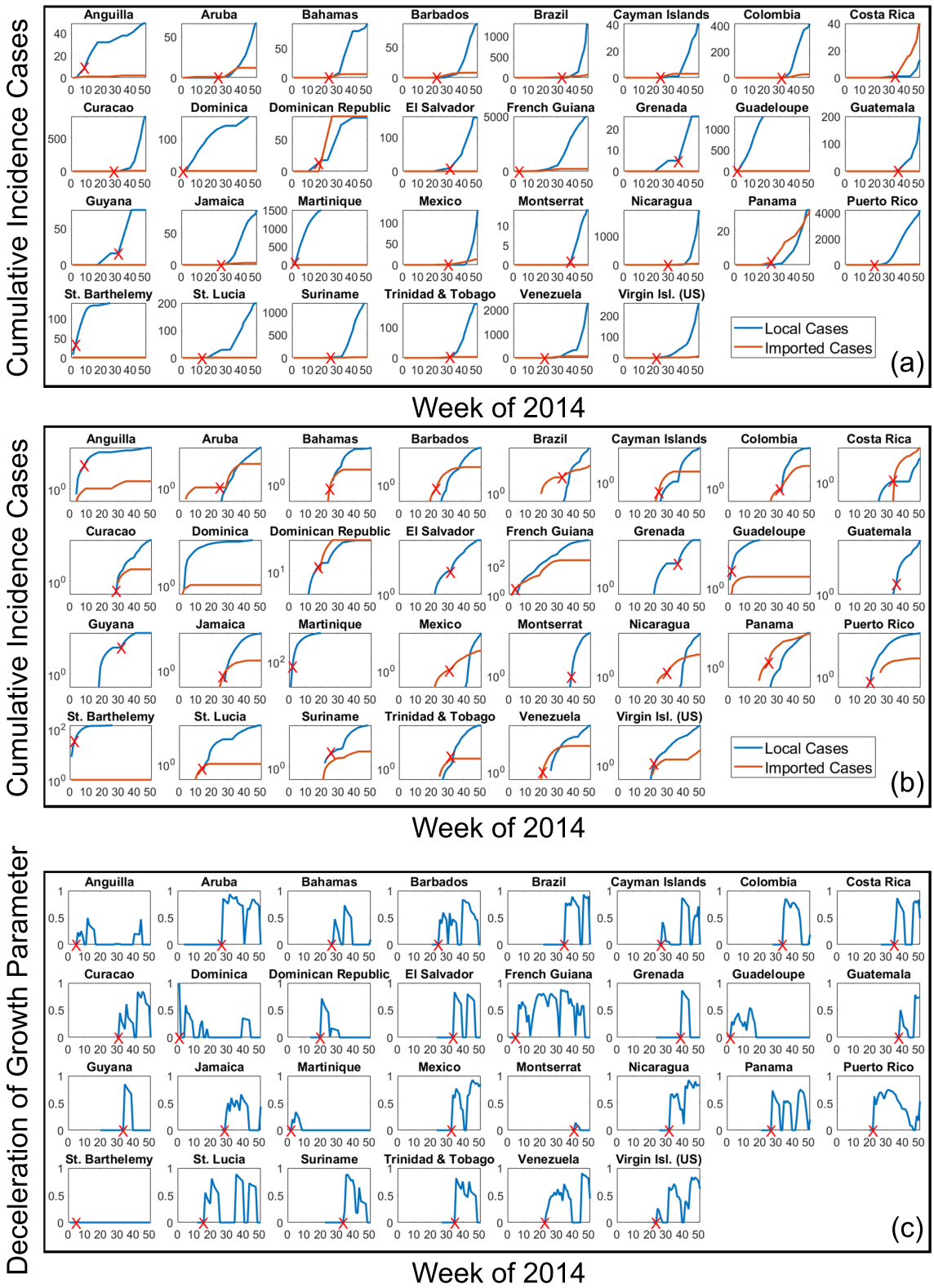

Figure 5: Analysis of chikungunya infection cases in the Western hemisphere. (a) The number of local and imported cases by week. The red 'X' indicates the establishment time of local transmission, as computed by the deceleration of growth parameter. (b) The same data as in (a) but on a log scale, to illustrate the early outbreak dynamics. (c) The evolution of the deceleration of growth parameter over time.

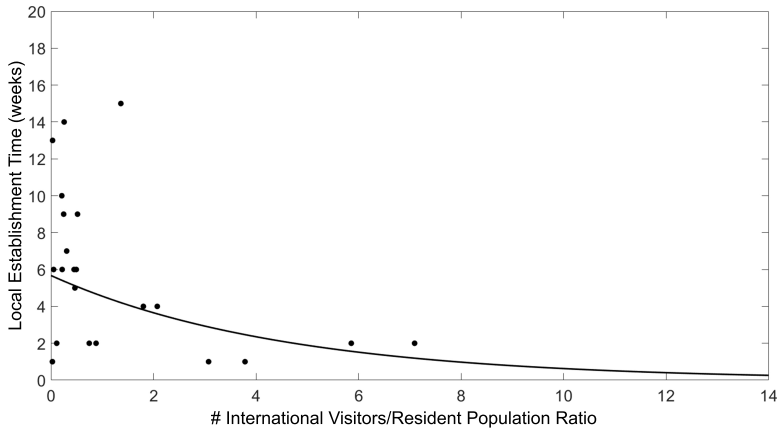

Figure 6: Relationship between tourism and the time to local establishment. The x-axis shows the total number of international visitors in 2014 divided by the resident population of each country. The y-axis is the time difference between the first case and the onset of local transmission in each country. The regression line is a log curve.

settlements on the rural outskirts of a population. In the small-world and triangulation models, which have more degree equitable distributions, the standard deviation in local establishment time is almost negligible. Therefore, the local establishment time is a useful diagnostic tool for predicting the rate at which an epidemic can establish itself across an entire metapopulation. This average local establishment time for an entire network is closely correlated to its average connectivity and clustering among its nodes, and would be straightforward to compute in practice.

We tested other epidemic models, such as the Susceptible-Exposed-Infected-Recovered (SEIR) model, to confirm that the results of this paper can be applied more generally to the field of epidemiology. The estimates of local establishment time in an SEIR model, for the scale-free, small-world, and triangulation networks, are qualitatively similar to the results shown in Figure 4, but with greater uncertainty due to the additional transition time between the exposed and infected states. The analytical model and the numerical simulations both treat the incidence curve as a continuous function and assume the human behavior is constant over time, both of which clearly do not apply to real-world epidemic scenarios. It is therefore not straightforward to make the same broad judgments about the empirical example in Section 3.3. However, it can still be inferred that countries with large quantities of human mobility, relative to their population, have shorter local establishment times. In future epidemic scenarios, the local establishment time within a given country can be estimated fairly reliably based on the properties of the disease (transmission and recovery rates). With additional data, it may also be possible to predict how travel advisories that reduce tourism would in turn increase the local establishment time.

4.1 Model Limitations

There are numerous ways in which the models described in this article can be expanded upon. This article does not consider the influence of environmental and socioeconomic factors on disease vector habitat suitability (Caminade et al., 2017). We propose the hypothesis that certain subpopulations can be high risks for epidemic outbreaks based on some combination of its network structure, its habitat suitability, and its capacity to respond to infection cases. Fortunately, simulating these varying factors is straightforward within the context of a metapopulation SIR model. One just needs to vary the quantities of β and μ across the network, in response to empirical factors such as temperature, elevation, population density, and access to medical services. Along similar lines, we can test the effects of government enforced travel restrictions on the local establishment time by lowering the quantity of the diffusion rate p accordingly.

In addition, this article only discusses the effects of the *external* network structure on the development of an epidemic within a node. This model assumes that human interaction within a subpopulation is homogeneous, that there is an equal probability per unit time that any random pair of humans will interact with each other. A more advanced model should consider heterogeneous distributions of population within each node. Possible solutions include modeling individual nodes as spatial lattices or nested networks. Population nodes can be quantified based on their external connectivity, clustering coefficient, and centrality as members of the larger population network, as well as their internal degree distribution, average clustering coefficient, and average centrality of their subnetwork structure (Chowell et al., 2017).

5 Conclusion

This article offers illumination to a poorly understood process that underlies epidemic outbreaks within a given population, and it is intended as a foundation into future research. Further analysis will discuss the effects on the local establishment time of heterogeneous population distributions, human behavior, and diseases whose parameters are spatially and temporally varying. It is also necessary to further explore the practical application of this analysis on predicting and mitigating future epidemic scenarios.

Our eventual goal is to develop an epidemic model in which the node, rather than the individual, is the unit of computation. This would allow for the extremely rapid and streamlined analysis of epidemic outbreaks on human population networks and inform the study of how the structure of these networks influences the spread of the epidemic. Understanding the time duration between the different stages of an epidemic within a subpopulation is an integral component of this analysis.

6 Acknowledgments

This project was funded by NSF CNH award #1824961 to AMG, BAL, and AS and USDA Hatch Project #ME021905 to AMG. Thank you to our collaborators Brian Allan, Sandra De Urioste-Stone, Andrew Mackay, and Shaowen Wang for their advice and guidance throughout the course of this project.

References

Adams, B. and D. D. Kapan (2009). Man Bites Mosquito: Understanding the Contribution of Human Movement to Vector-Borne Disease Dynamics. *PLoS ONE 4*(8), 6763. 111

Albert, R. and A. L. Barabási (2002, jun). Statistical mechanics of complex networks. *Reviews of Modern Physics 74*(1), 47–97. 118

Anderson, R. M. and R. M. May (1984, jan). Spatial, temporal, and genetic heterogeneity in host populations and the design of immunization programmes. *Mathematical Medicine and Biology 1*(3), 233–266. 112

Apolloni, A., C. Poletto, J. J. Ramasco, P. Jensen, and V. Colizza (2014). Metapopulation epidemic models with heterogeneous mixing and travel behaviour. *Theoretical Biology and Medical Modelling 11*(1). 113

Ball, F., T. Britton, and D. Sirl (2013, mar). A network with tunable clustering, degree correlation and degree distribution, and an epidemic thereon. *Journal of Mathematical Biology 66*(4-5), 979–1019. 112

Barrat, A., M. Barthelemy, R. Pastor-Satorras, and A. Vespignani (2003, mar). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences 101*(11), 3747–3752. 114

Benedict, M. Q., R. S. Levine, W. A. Hawley, and L. P. Lounibos (2007). Spread of the tiger: global risk of invasion by the mosquito Aedes albopictus. *Vector borne and zoonotic diseases (Larchmont, N.Y.) 7*(1), 76–85. 111

Briegel, H., I. Knüsel, and S. E. Timmermann (2001). Aedes aegypti: size, reserves, survival, and flight potential. *Journal of Vector Ecology 26*, 21–31. 114

Caminade, C., J. Turner, S. Metelmann, J. C. Hesson, M. S. C. Blagrove, T. Solomon, A. P. Morse, and M. Baylis (2017). Global risk model for vector-borne transmission of Zika virus reveals the role of El Niño 2015. *Proceedings of the National Academy of Sciences of the United States of America 114*(7), E1301–E1302. 123

Carnegie, N. B. (2018, jan). Effects of contact network structure on epidemic transmission trees: implications for data required to estimate network structure. *Statistics in Medicine 37*(2), 236–248. 112

Chowell, G., L. Sattenspiel, S. Bansal, and C. Viboud (2017). Mathematical models to characterize early epidemic growth: A Review. *Phys Life Rev 18*, 66–97. 114, 123

Chowell, G., C. Viboud, L. Simonsen, and S. M. Moghadas (2016, oct). Characterizing the reproduction number of epidemics with early subexponential growth dynamics. *Journal of The Royal Society Interface 13*(123), 20160659. 111

Cohen, M. L. (2000, aug). Changing patterns of infectious disease. *Nature 406*(6797), 762–767. 111

Colizza, V. and A. Vespignani (2008). Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *Journal of Theoretical Biology 251*(3), 450–467. 114, 115

Cross, P. C., J. O. Lloyd-Smith, P. L. F. Johnson, and W. M. Getz (2005). Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. *Ecology Letters 8*(6), 587–595. 116

Keeling, M. J. and P. Rohani (2007). *Modeling Infectious Diseases in Humans and Animals* (1 ed.). Princeton University Press. 114

Lee, D. T. and B. J. Schachter (1980, jun). Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer & Information Sciences 9*(3), 219–242. 118

Lipsitch, M., C. A. Donnelly, C. Fraser, I. M. Blake, A. Cori, I. Dorigatti, N. M. Ferguson, T. Garske, H. L. Mills, S. Riley, M. D. Van Kerkhove, and M. A. Hernán (2015, jul). Potential Biases in Estimating Absolute and Relative Case-Fatality Risks during Outbreaks. *PLOS Neglected Tropical Diseases 9*(7), e0003846. 115

Lounibos, L. P. (2002, jan). Invasions by Insect Vectors of Human Disease. *Annual Review of Entomology 47*(1), 233–266. 111

Moore, S. M., Q. A. ten Bosch, A. S. Siraj, K. J. Soda, G. España, A. Campo, S. Gómez, D. Salas, B. Raybaud, E. Wenger, P. Welkhoff, and T. A. Perkins (2018). Local and regional dynamics of chikungunya virus transmission in Colombia: The role of mismatched spatial heterogeneity. *BMC Medicine 16*(1), 152. 114

Nah, K., K. Mizumoto, Y. Miyamatsu, Y. Yasuda, R. Kinoshita, and H. Nishiura (2016). Estimating risks of importation and local transmission of Zika virus infection. *PeerJ 2016*(4). 111

PAHO (2015). PAHO | Chikungunya. 115

Pastor-Satorras, R., C. Castellano, P. Van Mieghem, and A. Vespignani (2015). Epidemic processes in complex networks. *Reviews of Modern Physics 87*(3). 111, 113

Ray, C., M. Hoopes, I. Hanski, and M. E. Gilpin (1997). Metapopulation Biology: Ecology, Genetics, and Evolution. *Ecology 78*(7), 2270. 112

Riou, J., C. Poletto, and P. Y. Boëlle (2017, jun). A comparative analysis of Chikungunya and Zika transmission. *Epidemics 19*, 43–52. 111

Staples, J., R. Breiman, and A. Powers (2009, sep). Chikungunya Fever: An Epidemiological Review of a Re-Emerging Infectious Disease. *Clinical Infectious Diseases 49*(6), 942–948. 114

Steele, L., E. Orefuwa, and P. Dickmann (2016, dec). Drivers of earlier infectious disease outbreak detection: a systematic literature review. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases 53*, 15–20. 114

Tomasello, D. and P. Schlagenhauf (2013, sep). Chikungunya and dengue autochthonous cases in Europe, 2007–2012. *Travel Medicine and Infectious Disease 11*(5), 274–284. 114

Trapman, P., F. Ball, J.-S. Dhersin, V. C. Tran, J. Wallinga, and T. Britton (2016). Inferring R0 in emerging epidemics-the effect of common population structure is small. *Journal of the Royal Society, Interface 13*(121). 112

Valdano, E., M. R. Fiorentin, C. Poletto, and V. Colizza (2018). Epidemic threshold in continuous-time evolving networks. *Physical Review Letters 120*(6). 111

Viboud, C., O. N. Bjørnstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell (2006). Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science 312*(5772), 447–451. 111

Watts, D. J. and S. H. Strogatz (1998, jun). Collective dynamics of 'small-world' networks. *Nature 393*(6684), 440–2. 118

Weaver, S. C. and N. L. Forrester (2015, aug). Chikungunya: Evolutionary history and recent epidemic spread. *Antiviral Research 120*, 32–39. 114

World Bank (2019). World Development Indicators | DataBank. 115

Zhang, J. P. and Z. Jin (2011). The analysis of an epidemic model on networks. *Applied Mathematics and Computation 217*(17), 7053–7064. 111