

Modeling outbreak data: Analysis of a 2012 Ebola virus disease epidemic in DRC

Boseung Choi^{*}, Sydney Busch[†], Dieudonné Kazadi^{††‡‡}, Benoit Ilunga^{‡‡}, Emile Okitolonda^{††}, Yi Dai[‡], Robert Lumpkin[§], Omar Saucedo[¶], Wasiur R. KhudaBukhsh^{‡¶}, Joseph Tien[§], Marcel Yotebieng^{||}, Eben Kenah[‡], Grzegorz A. Rempala^{‡§¶**}

^{*}Department of National Statistics, Korea University Sejong Campus
Sejong, Republic of Korea

[†]Department of Mathematics, Augsburg College
Minneapolis, MN, USA

^{††}Ministry of Health, Democratic Republic of the Congo

^{‡‡}School of Public Health, University of Kinshasa
Kinshasa, Democratic Republic of the Congo

[‡]Division of Biostatistics, College of Public Health

[§]Department of Mathematics

[¶]Mathematical Biosciences Institute

^{||}Division of Epidemiology, College of Public Health
The Ohio State University, Columbus, OH, USA

^{**}Corresponding author: rempala.3@osu.edu

Received: 8 September 2019, accepted: 3 October 2019, published: 15 October 2019

Abstract—We describe two approaches to modeling data from a small to moderate-sized epidemic outbreak. The first approach is based on a branching process approximation and direct analysis of the transmission network, whereas the second one is based on a survival model derived from the classical SIR equations with no explicit transmission information. We compare these approaches using data from a 2012 outbreak of Ebola virus disease caused by *Bundibugyo ebolavirus* in city of Isiro, Democratic Republic of the Congo. The branching process model allows for a direct comparison of

disease transmission across different environments, such as the general community or the Ebola treatment unit. However, the survival model appears to yield parameter estimates with more accuracy and better precision in some circumstances.

Keywords—parameter estimation; branching process; Markov Chain Monte-Carlo methods; survival dynamical system;

I. INTRODUCTION

On August 1, 2018, the Ministry of Health of the Democratic Republic of the Congo (DRC)

Copyright: © 2019 Choi et al. This article is distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Boseung Choi, Sydney Busch, Dieudonné Kazadi, Benoit Ilunga, Emile Okitolonda, Yi Dai, Robert Lumpkin, Omar Saucedo, Wasiur R. KhudaBukhsh, Joseph Tien, Marcel Yotebieng, Eben Kenah, Grzegorz A. Rempala, Modeling outbreak data: Analysis of a 2012 Ebola virus disease epidemic in DRC, *Biomath* 8 (2019), 1910037, <http://dx.doi.org/10.11145/j.biomath.2019.10.037>

reported an outbreak of Ebola virus disease (EVD) in North Kivu Province. At the time of writing about one year later, confirmed and probable cases have been reported in nine health zones of North Kivu and Ituri provinces (including the provincial capital city of Goma), threatening further spread of the epidemic into neighboring provinces and the countries of Uganda and Rwanda. The current outbreak area is roughly 780 miles away from Equateur province, where an earlier Ebola outbreak was reported in May 2018. This persistent reoccurrence of Ebola in the DRC as well as elsewhere in Africa is a reminder that another large pandemic like the 2013–2016 West African Ebola epidemic remains possible. To control future outbreaks and to better understand patterns of transmission in households and at health care facilities, it is essential to carefully analyze well-documented historic data from past outbreaks.

The current paper is concerned with modeling data from a small 2012 Ebola virus disease (EVD) outbreak caused by *Bundibugyo ebolavirus* (BDBV) in the Isiro municipality in DRC. The interesting feature of this dataset is that it includes partial contact information on Ebola cases treated either in the community or in healthcare facilities, which allows for network-based inference. Despite the fact that such inference has been an extremely active area of research in the past 20 years [1]–[5], there have been relatively few well documented historical datasets from real epidemics. For our purpose of analyzing a relatively small network, we here apply the *edge-based* approach of Miller and Volz [15]–[17] and compare it with the recently proposed simple non-network *survival dynamical system* model [18].

The paper is organized as follows. In the remainder of this section we give some basic background information on the Ebola virus and describe the Isiro outbreak dataset to be analyzed. In Sections 2 and 3, we outline the proposed statistical models, use them to analyze the Isiro outbreak data, and describe the results. In Section 4, we offer some brief concluding remarks.

A. Ebola virus disease

EVD in humans is a severe hemorrhagic fever caused by four of the five viruses of the genus *Ebolavirus* in the viral family *Filoviridae* [19]. The virus was initially characterized in 1976 during an outbreak in DRC (known then as Zaire) and has subsequently caused at least 26 additional outbreaks in human populations [20], [21]. Since 2000, there have been 8 recorded outbreaks in the DRC. The ongoing outbreak in the North Kivu province is the most serious to date, and Maganga et al. [22] describe an earlier serious outbreak that occurred in the Boende region of Équateur Province in 2014. Data is available in the literature for at least 4 out of these 8 outbreaks, although these data are not considered fully reliable because the majority of the outbreaks occurred in remote areas. Although EVD outbreaks in DRC have been limited to fewer than 500 cases to date, the potential for dangerous spread across the African continent was demonstrated by the 2013–2016 West African Ebola epidemic, which resulted in more than 28,600 cases and 11,313 deaths across ten countries [23].

The introduction of the virus into human communities is likely the result of sporadic zoonotic events. Several species of fruit bats native to areas endemic to Ebola have been implicated as the natural reservoir for the disease [19]. Upon infection, the virus will typically incubate for a period of two to 21 days [23]. After this period, the typical clinical presentation is a mix of severe symptoms that may include fever, nausea, diarrhea, vomiting, chest pain, dyspnea, cough, ocular edema, hypotension, conjunctivitis, headaches, coma, and hemorrhaging [24].

Human-to-human Ebola transmission is believed to occur via close contact with infected bodily fluids and, therefore, individuals such as family members of cases and healthcare workers have significantly increased risk of infection [25]–[28]. In addition, viral load (which changes throughout the course of illness and is at the highest level immediately after death) is known to impact the probability of transmission [29], [30]. Transmis-

sion dynamics are further complicated by the fact that an infectious individual's contacts may vary throughout disease progression. Accounting for these dynamics through contact tracing and network analysis is vital to understanding the disease, as studies of previous outbreaks have noted disproportionately high rates of infection among women, healthcare workers, and family members [23], [25], [27], [31].

Practical strategies have been implemented to reduce the risk of transmission, including barrier nursing methods, safe burial practices, and the creation of isolation units within treatment centers [31], [32]. Nevertheless, the scope of the 2013-2016 West African Ebola epidemic suggests that a deeper quantitative understanding of these dynamics might be needed to control transmission more effectively. While an effective vaccine is now available [33], the question of how and when to most effectively use available resources during an ongoing epidemic remains in need of further quantitative study.

Prior to the West African Ebola epidemic, only a handful of mathematical models for Ebola had been studied [34]–[37]. Chowell et al. [34] and Lekone and Finkenstadt [35] used an SEIR framework to determine the effect of interventions on the 1995 DRC and 2000 Uganda outbreaks. Legrand et al. [36] formulated a stochastic compartmental model that accounted for transmission in several epidemiological settings by introducing compartments for hospitalized individuals and dead Ebola cases, who can transmit the disease during funerals.

The West African epidemic highlighted the critical need for a better understanding of the Ebola transmission dynamics and potential control measures. Consequently, there has been an outpouring of Ebola models including deterministic compartmental models [38]–[40], stochastic models [41]–[44] and multi-type branching process models [45]. Many of these adapted the SEIHRF framework, which includes hospital and funeral compartments [26], [46]–[48].

Recent work has considered spatial aspects of

Ebola virus transmission and the effect of clustering in the population. The spatiotemporal spread of Ebola has been studied with a county-level multi-patch model employing mobile data [41], with spatial individual-based models for international spread [43], and transmission between households [44]. Scarpino et al. [26] used a phylodynamic model to reconstruct chains of transmission for cases occurring in Sierra Leone in June 2014. Their fit of an SEIR model with the Rand-style [1] pair approximation gave evidence for the presence of clustering within the population. The same pair approximation was employed by Wells et al. [46] in their investigation of the effectiveness of case isolation and ring vaccination. Their model had 10 compartments and required 65 equations after closure at the level of pairs.

B. Isiro EVD outbreak data

Isiro is a municipality in the north-east part of DRC that is the capital of Haut-Uele district. It is situated between equatorial forest and savannah. The Isiro dataset contains information about the 2012 EVD outbreak caused by *Bundibugyo ebolavirus* (BDBV). In total, there were 62 cases of infection listed as either probable or confirmed with 52 of them having proper clinical information allowing for a more detailed study. As shown in Figure 1(a), these were divided into community cases and Ebola treatment center (ETC) cases based on the source of the infecting contact. Among community cases, there was overrepresentation of females (85.3%) and of individuals aged 15-54 years (82.4%). The mean duration of EVD was 18 days, and the mean incubation period was 11.3 days [49].

We were able to obtain additional information concerning contacts for most individual cases from the DRC Ministry of Health. However, unlike traditional contact tracing, these additional records contained a list of potentially infecting individuals for each case. Using this information, we were able to track the likely number of people each case infected as well as the overall contact network (for most of the EVD cases) in both the community and the ETC. To refine the transmission network,

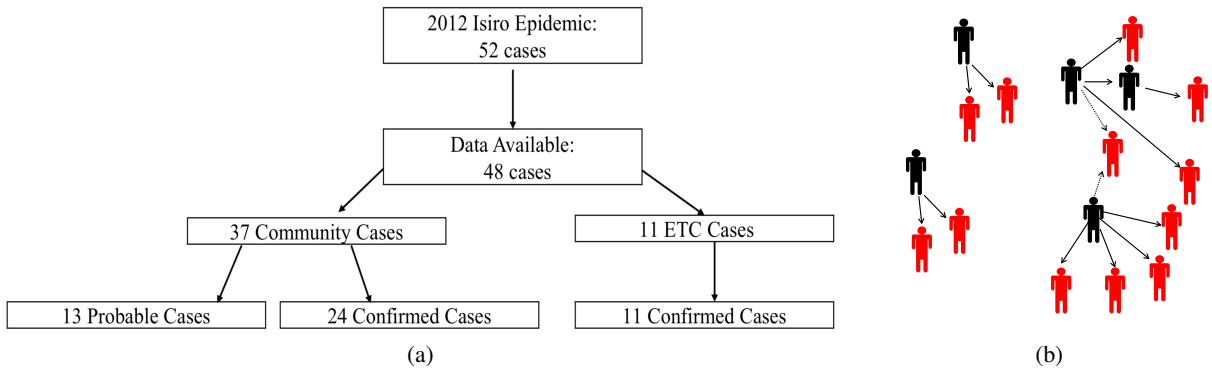


Fig. 1: **2012 Isiro EVD data and model.** Panel (a): Summary of available Isiro cases used in current analysis of the transmission dynamics in the community and ETC. This data is a subset of 52 cases described in [49]. Panel (b): Example of transmission data reconstructed from the Isiro outbreak files. Dark figures represent primary cases and secondary cases who infected others. All others represent infected who did not transmit. All cases of transmission ambiguity (multiple in-arrows) were resolved uniformly at random.

we used occupation and other socioeconomic information as available. Of the 52 documented infections in the Isiro epidemic, clinical documentation of contacts among 48 cases (17 probable, 31 confirmed) could be retrieved. Out of these, there were 37 community cases¹ (13 suspected or probable and 24 confirmed) who never reached the ETC. The 11 cases that reached the ETC were all confirmed as EVD [49].

II. STATISTICAL MODELS

In this paper, we consider two different statistical models for the Isiro EVD outbreak. The first one is based on a branching process approximation of the virus transmission network, which appears especially appropriate for directly comparing outbreak parameters in different environments such as the village community and the ETC. The second model is based on a survival analysis approach that assumes homogenous contact patterns among susceptibles and infectives. This assumption is more likely to be appropriate among community cases than ETC cases.

¹We note here a slight discrepancy with [49] where only 34 cases were classified as community.

A. Branching model

1) *Primary cases:* For a primary (or index) case i represented by a node of degree d_i , the distribution of the number of secondary infections created by i (say, X_i) conditionally on the degree and the infection period (say, t_i) is given by

$$P(X_i = x_i | t_i, d_i) = \binom{d_i}{x_i} p_{t_i}^{x_i} (1 - p_{t_i})^{d_i - x_i}. \quad (1)$$

We see therefore that the distribution function of X_i is binomial with d_i trials and the probability of success p_{t_i} . Under the assumption that infectious contacts follow a Poisson process with rate β , the probability of a successful infection of a given neighbor by case i before time t_i is

$$p_{t_i} = 1 - e^{-\beta t_i}.$$

The infectious period (i.e., time from symptom onset to removal) for each case is assumed to follow an exponential distribution with rate γ , so its density is

$$f(t_i) = \gamma e^{-\gamma t_i}. \quad (2)$$

For the i -th index case the joint conditional probability distribution of (x_i, t_i) given d_i is the product

of (1) and (2):

$$\begin{aligned} \pi_{d_i}(x_i, t_i) &= P(X_i = x_i | t_i, d_i) f(t_i) \\ &= \binom{d_i}{x_i} \gamma (1 - e^{-\beta t_i})^{x_i} e^{-(\beta(d_i - x_i) + \gamma)t_i}. \end{aligned}$$

Without any prior information on the degree d_i of the i -th index case, the unconditional joint probability distribution of the number of infections and the infection period is then given by

$$f(x_i, t_i) = \sum_{d_i \geq x_i} q_{d_i} \pi_{d_i}(x_i, t_i) \quad (3)$$

where q_d is the probability of degree d . Here, we consider the Poisson distribution with mean λ , which appears to fit well the transmission network for the Isiro epidemic. With this degree distribution, the final form of (3) admits the following closed-form expression parametrized by the triple $\theta = (\beta, \gamma, \lambda) \in \mathbb{R}_{>0}^3$:

$$\begin{aligned} f_\theta(x_i, t_i) &= \sum_{d_i \geq x_i} \frac{\lambda^{d_i} e^{-\lambda}}{d_i!} \binom{d_i}{x_i} \gamma (1 - e^{-\beta t_i})^{x_i} e^{-(\beta(d_i - x_i) + \gamma)t_i} \\ &= \frac{[\lambda(1 - e^{-\beta t_i})]^{x_i}}{x_i!} e^{-\lambda(1 - e^{-\beta t_i})} \gamma e^{-\gamma t_i}. \end{aligned} \quad (4)$$

2) *Secondary cases*: Consider now the joint distribution of the number of infections and the infection period for a secondary (i.e., non-index) case. For tractability, we assume the random network follows the configuration model (CM), so edges are formed uniformly at random (excluding multiple edges and self-loops) given the degrees of all nodes. For such networks, we define a *secondary case* as an individual (node) to whom the infection was successfully transmitted (see also [50]). Note that since by definition the secondary case has one infecting neighbor, its degree available for further infection decreases by one. This decreased degree is often referred to as the *excess degree* in the literature.

For secondary cases, we can modify the primary case model by replacing the degree with the excess degree. Let q'_{d_j} denote the excess degree probability for a secondary case with degree $d_j > 0$, and let

$\mu \in (0, \infty)$ be the mean of the degree distribution. Then it is known ([51]) that for the CM network

$$q'_{d_j} = \frac{d_j q_{d_j}}{\mu}.$$

In the case of the Poisson degree distribution, $\lambda = \mu$ and it is easy to check that

$$q'_{d_j} = q_{d_j - 1}. \quad (5)$$

Let x'_j and t'_j be the number of infections and the infection period for the secondary case j . Equations (3) and (5) give us

$$\begin{aligned} f(x'_j, t'_j) &= \sum_{d_j > x'_j} q'_{d_j} \pi_{d_j - 1}(x'_j, t'_j) \\ &= \sum_{d_j > x'_j} q_{d_j - 1} \pi_{d_j - 1}(x'_j, t'_j) \\ &= f_\theta(x'_j, t'_j), \end{aligned} \quad (6)$$

where f_θ is given by (4).

For known $\theta = (\beta, \gamma, \lambda)$, we can calculate the key characteristic of an epidemic known as the *basic reproduction number* (\mathcal{R}_0). It may be interpreted as the average number of secondary infections caused by a primary case. For an arbitrary degree distribution the definition of \mathcal{R}_0 for CM graph is given, for instance, in [52]. For the Poisson degree distribution with exponential infectious periods, we have simply

$$\mathcal{R}_0 = \frac{\beta \lambda}{\beta + \gamma}. \quad (7)$$

3) *Likelihood and estimation*: Suppose the observed numbers of infections and infectious periods for m primary cases are $\mathbf{x} = \{x_1, x_2, \dots, x_m\}$ and $\mathbf{t} = \{t_1, t_2, \dots, t_m\}$, and those for n secondary cases are $\mathbf{x}' = \{x'_1, x'_2, \dots, x'_n\}$ and $\mathbf{t}' = \{t'_1, t'_2, \dots, t'_n\}$. Recall that $\theta = (\beta, \gamma, \lambda)$ is the vector of parameters that need to be estimated. Here β is the transmission rate of infection, and γ^{-1} is the mean infectious period, and λ is the mean degree. For the observed data, the likelihood function for θ can be constructed using (4) and (6):

$$\mathcal{L}(\theta | \mathbf{x}, \mathbf{t}, \mathbf{x}', \mathbf{t}') \propto \prod_{i=1}^m f_\theta(x_i, t_i) \prod_{j=1}^n f_\theta(x'_j, t'_j).$$

We could calculate the maximum likelihood estimates (MLEs) of θ by maximizing the expression above using numerical optimization. However, for better stability and reproducibility of results, we will use a Bayesian Markov Chain Monte Carlo (MCMC) procedure. We assign independent gamma prior distributions for the components of θ . Specifically, let

$$\begin{aligned} \pi(\beta) &\sim \Gamma(a_\beta, b_\beta) \\ \pi(\gamma) &\sim \Gamma(a_\gamma, b_\gamma) \\ \pi(\lambda) &\sim \Gamma(a_\lambda, b_\lambda). \end{aligned} \tag{8}$$

Given the data \mathbf{x} , \mathbf{t} , \mathbf{x}' , and \mathbf{t}' , the Bayesian estimate of θ is the mean of the joint posterior distribution

$$\mathcal{L}^p(\theta|\mathbf{x}, \mathbf{t}, \mathbf{x}', \mathbf{t}') \propto \mathcal{L}(\theta|\mathbf{x}, \mathbf{t}, \mathbf{x}', \mathbf{t}')\pi(\beta)\pi(\gamma)\pi(\lambda). \tag{9}$$

This mean may be approximated with an empirical average from the following converged MCMC sampler:

Algorithm 1 MCMC posterior sampler for the branching process model

- 1: Initialize θ with $\theta_0 = (\beta_0, \gamma_0, \lambda_0)$.
 - 2: Sample β via a Metropolis-Hastings step [53] with truncated normal proposal for the target conditional distribution of β given $\mathbf{x}, \mathbf{t}, \mathbf{x}', \mathbf{t}', \lambda$.
 - 3: Sample γ directly from its conditional distribution

$$\gamma|\mathbf{x}, \mathbf{t}, \mathbf{x}', \mathbf{t}' \propto \gamma^{m+n+a_\gamma-1} e^{-\gamma(\sum t_i + \sum t'_j + b_\gamma)}.$$
 - 4: Sample λ via another Metropolis-Hastings step (similar to Step 2) for the target conditional distribution of λ given $\mathbf{x}, \mathbf{t}, \mathbf{x}', \mathbf{t}', \beta$.
 - 5: Return to Step 2 and repeat until convergence.
-

B. Survival model

Under the survival model, it is assumed that the population at risk (which is possibly much smaller than the set of all initially susceptible individuals) interacts homogeneously according to the Kermack-McKendrick model (see below). Hence

it is problematic to apply this model in the case of an ETC where the homogenous mixing is likely not satisfied. Consequently, we apply this model only to the community outbreak data. In what follows it is convenient to write the usual Kermack-McKendrick model in the following integral form:

$$\begin{aligned} s_t &= \exp\left(-\beta \int_0^t \iota_u du\right) = \exp(-\mathcal{R}_0 r_t) \\ \iota_t &= \rho e^{-\gamma t} - \int_0^t s_u e^{-\gamma(t-u)} du \\ r_t &= \gamma \int_0^t \iota_u du \end{aligned}$$

where $\rho = \iota_0$, and $\mathcal{R}_0 = \beta/\gamma$. By interpreting the strictly decreasing function s_t as an *improper survival function*, it follows [18] that the conditional density of infection time is given by

$$f_\tau(t) = -\dot{s}_t/\tau \tag{10}$$

where $\tau = \lim_{t \rightarrow \infty} r_t < 1$ is the final size of the epidemic [18], which is the unique solution of

$$1 - \tau = e^{-\mathcal{R}_0(\tau+\rho)}. \tag{11}$$

Thus, for a collection of n individuals initially at risk, out of which k are infected at respective times $t_1 < \dots < t_k < T$ (where $T < \infty$ is the maximum follow-up time), we have the following log-likelihood function for infection times

$$l_I(t_1, \dots, t_k|\theta, n) = (n-k) \log s_T + \sum_{i=1}^k \log f_\tau(t_i).$$

Note that this likelihood is conditional on the number of individuals at risk, which is often unknown. However, given the number of infections (k) by time T and under the assumption of independence of infection times (see [18] for discussion), we may consider n as the realization of the negative binomial distribution: $n \sim \text{NegBinom}(k, \tau)$. Denote by w_i the i th individual's *removal time*, defined as minimum of the times of individual's recovery, hospitalization, or death. Assuming r recoveries given k infections, the log-likelihood is

$$\begin{aligned} l_R(w_1, \dots, w_r|\theta, k) \\ = (k - r) \log H_\gamma(T) + \sum_{i=1}^r \log h_\gamma(w_i), \end{aligned}$$

where $H_\gamma(\cdot)$ is the survival function and $h_\gamma(\cdot)$ is the PDF of an exponential distribution with rate γ . The complete log-likelihood is then the sum of l_I and l_R [18]:

$$l_0(t_1, \dots, t_k, w_1, \dots, w_r | \theta, n) = l_I(t_1, \dots, t_k | \theta, n) + l_R(w_1, \dots, w_r | \theta, k).$$

Under this survival model, we may estimate the model parameters and the size of the population at risk n using another Bayesian MCMC. The model parameters are now $\theta' = (\beta, \gamma, \rho)$, because τ is fully determined by θ' via (11). The prior distributions for β and γ are of the form (8) and the prior distribution for ρ is taken as Uniform(0, 1). All parameters are assumed independent a priori. The MCMC algorithm used to obtain the posterior sample of the parameters is similar to that in previous section.

Algorithm 2 MCMC posterior sampler for the survival model

- 1: Initialize $\theta' = (\beta, \gamma, \rho)$ from the prior distribution and set $n = k$.
 - 2: Perform a Metropolis-Hastings step (using the truncated normal proposal) for the target conditional distribution of θ' given n using the complete log-likelihood $\ell_0 = \ell_I + \ell_R$.
 - 3: Calculate τ based on the current value of θ' as the solution to final size equation (11).
 - 4: Sample the conditional distribution of n given θ' by drawing $n \sim \text{NegBinom}(k, \tau)$.
 - 5: Return to Step 2 and repeat until convergence.
-

III. DATA ANALYSIS

The analysis was conducted separately for the two models (branching process and survival) based on the 48 available cases of EVD described in Section 1. Parameter estimates were obtained using the MCMC algorithms for each model described in Section 2. For the branching process model, we separately analyzed the community and ETC outbreaks. For the survival model, we only analyzed the community outbreak.

A. Branching process model

To perform the separate analyses of the community and ETC outbreaks, the contact and infection data were partitioned into two subsets depending on the location of the infective contact (community or ETC). Although the same uninfected individuals were allowed to be in both outbreak networks, all the EVD cases were assigned either to the community or to the ETC. When reconstructing the transmission network, all ambiguous contact tracing was resolved uniformly at random as shown in Figure 1(b). For some individuals, the complete infection period was unknown and needed to be imputed. All such imputations were based on the density (2) and performed between steps 3 and 4 in MCMC algorithm.

For estimating β and γ , we assigned non-informative gamma prior distributions with location and rate parameters of 0.001. However, due to limited contact information, we assigned a relatively informative prior to λ with location parameter of 6 and a rate parameter of 1. This assignment was based on the empirical mean of the primary and secondary cases in the dataset. Finally, for the Metropolis-Hastings steps in the MCMC sampler algorithm we used the truncated normal distribution as proposal distribution and tuned its standard deviation to achieve an acceptance ratio of 44% as recommended in [54]. The final results of the MCMC were based on 55,000 iterations of the sampler with first 5,000 iterations removed as “burn-in”. The posterior samples were thinned by keeping only the results from every 10th iteration, resulting in a final set of 5,000 posterior samples that were used to estimate the parameters and calculate approximate posterior credible intervals. The convergence of the MCMC algorithm was diagnosed based on the R statistic and trace and autocorrelation (ACF) plots. To conserve space, these plots are not shown here.

Table I summarizes the results of branching process analyses of community and ETC epidemics. For each parameter, the posterior mean, standard deviation, and 95% credible interval (CI) are provided. We note that the comparisons be-

TABLE I: Results under the EVD branching process model for the community and ETC outbreaks

	Community infections			ETC infections		
	Mean	Std Dev	95% CI	Mean	Std Dev	95% CI
β	0.0741	0.0389	(0.0331, 0.1806)	0.0387	0.0182	(0.0152, 0.0851)
γ	0.1936	0.0397	(0.1254, 0.2811)	0.2205	0.0634	(0.1170, 0.3605)
λ	5.4460	1.4460	(2.9690, 8.6310)	5.9030	1.4200	(3.4200, 8.9820)
\mathcal{R}_0	1.3730	0.2951	(0.8510, 2.0230)	0.8592	0.3200	(0.3700, 1.6040)

TABLE II: Results under the EVD survival model for the community outbreak only

	Mean	Std Dev	95% CI
β	0.1964	0.0324	(0.1403, 0.2626)
γ	0.1774	0.0296	(0.1258, 0.2381)
ρ	0.0039	0.0017	(0.0017, 0.0079)
n	163.20	35.35	(113.00, 252.00)
\mathcal{R}_0	1.1080	0.0316	(1.0570, 1.1650)

tween parameters in the two epidemics may be conducted informally by comparing their respective CIs bounds. If a particular parameter’s CI bounds for the community are contained within the respective CI bounds for the ETC, or vice-versa, one would consider the corresponding posterior distributions as statistically (i.e., for given data) equal. For β , which represents the transmission rate of Ebola virus, the posterior mean for household infection was approximately 0.0741—about twice as large as ETC infection rate—so the two posterior distributions may be considered statistically different. This is in contrast with the parameter γ , which represents the reciprocal of the mean infectious period, for which the estimated respective posterior means of 0.1936 and 0.2205 for the community and ETC were not found to be statistically different based on their respective 95% credible bounds. The parameter λ represents the average degree of the degree distribution and its posterior mean in the ETC is slightly (but not significantly) larger than its posterior mean in the community. This may reflect additional contacts of the individuals at the ETC with patients, visitors, and ETC staff. Finally, the posterior means of the basic reproduction numbers \mathcal{R}_0 for the community and ETC outbreaks were found to be significantly different at 1.373 and 0.8592, respectively. As

expected, the posterior mean of \mathcal{R}_0 for the community outbreak is higher than one for the ETC. In both settings, \mathcal{R}_0 was calculated according to equation (7). However, we also note that the 95% CI bounds for both posterior distributions are quite wide, indicating a lack of precision in the branching process model.

B. Survival model

Under the survival model outlined in Section II-B, the analysis simplifies in that we are no longer concerned with estimating the network average degree λ . Hence our model parameters are $\theta' = (\beta, \gamma, \rho)$ as well as the size of the population at risk (or effective population size) denoted by n . In this model, $\mathcal{R}_0 = \beta/\gamma$. The results of the analysis for the 37 community cases are presented in Table II. We note that the estimates of the parameters $(\beta, \gamma, \mathcal{R}_0)$ under the survival model can be compared with the estimates of $(\beta\lambda, \gamma, \mathcal{R}_0)$ under branching process model. Despite considerable differences in their respective posterior mean values, the posterior distributions of all three parameters are statistically equivalent based on the respective 95% CI bounds. This underscores the lack of precision of the estimates based on the branching process model.

C. Model validation

To perform our model validation and fit assessment, we compared the distributions of 5,000 samples from the posterior distributions of the final outbreak size obtained from our branching process analysis to the values of 37 and 11 observed in the community and ETC outbreaks respectively. Figure 2 presents the histograms of the posterior size distributions for the respective final sizes. The vertical lines are plotted for reference at the observed outbreak values of 37 and 11. The histogram plots show that the observed values are close to the modes of the respective posterior distributions for both models, indicating adequate model fit [50]. The last panel of Figure 2 presents a comparison of observed depletion of susceptibles over the course of the community EVD outbreak with that predicted by the survival model. Both the observed and predicted curves are initiated on day one at the estimated mean of total population at risk. The two curves are very close to each other, indicating good fit of the survival model.

IV. SUMMARY AND CONCLUSIONS

We presented two statistical models for analyzing patterns of EVD transmission in a small community. The branching process model was based on partial contact network tracing data, whereas survival model used an aggregate (network-free) approach. The two models allowed for a more detailed analysis of the 2012 Isiro EVD epidemic data that was summarized in [49].

The branching process model was based on a configuration model random graph with a Poisson degree distribution, and it explicitly described the direct and indirect contacts of the EVD cases in the community and at the ETC. In particular, the model made it possible to derive and directly compare the characteristics of the EVD Isiro outbreaks at these two different locations. Although the comparison provided some evidence of the usefulness of the ETC in controlling EVD outbreaks, the analysis also suggests that the type of basic contact tracing performed in Isiro may not be sufficient to provide precise estimates of the

epidemic force via the basic reproduction number (\mathcal{R}_0).

The survival model was derived from the standard SIR equations. Since this model did not require contact tracing or estimation of mean degree, it did not suffer from the same problem of low precision as the branching process model. In fact, the estimate of \mathcal{R}_0 provided by the survival model was more precise and likely also more accurate (based on a resampling analysis not shown here) than the values obtained from the branching process model. However, the drawback of the survival model was that it could not be used for comparison of transmission in the community and the ETC because the ETC was unlikely to satisfy the homogeneous mixing assumption.

It appears that an effective approach to modeling the type of outbreak dynamics described by the Isiro data might be to combine the two models presented here, so as to retain the precision of the survival one but also incorporate the transmission network information. This will be likely the focus of our future research.

ACKNOWLEDGMENT

This research was partially funded by the US National Science Foundation and the US National Institutes of Health under grants DMS1853587, U54 GM111274 and R01 AI116770. The funds were also received from the Mershon Center for International Security Studies at OSU, the Korea University and the Ohio Five-OSU Summer Undergraduate Research Experience (SURE) Fund. The authors would like to especially thank the Mathematical Biosciences Institute (MBI) at OSU for providing additional resources and logistical help. MBI receives funds from the National Science Foundation under grants DMS1440386 and DMS1757423.

REFERENCES

- [1] H. Andersson and T. Britton, *Stochastic epidemic models and their statistical analysis*. Springer Science & Business Media, 2012, vol. 151.
- [2] D. Brockmann and D. Helbing, "The hidden geometry of complex, network-driven contagion phenomena," *Science*, vol. 342, no. 6164, pp. 1337–1342, 2013.

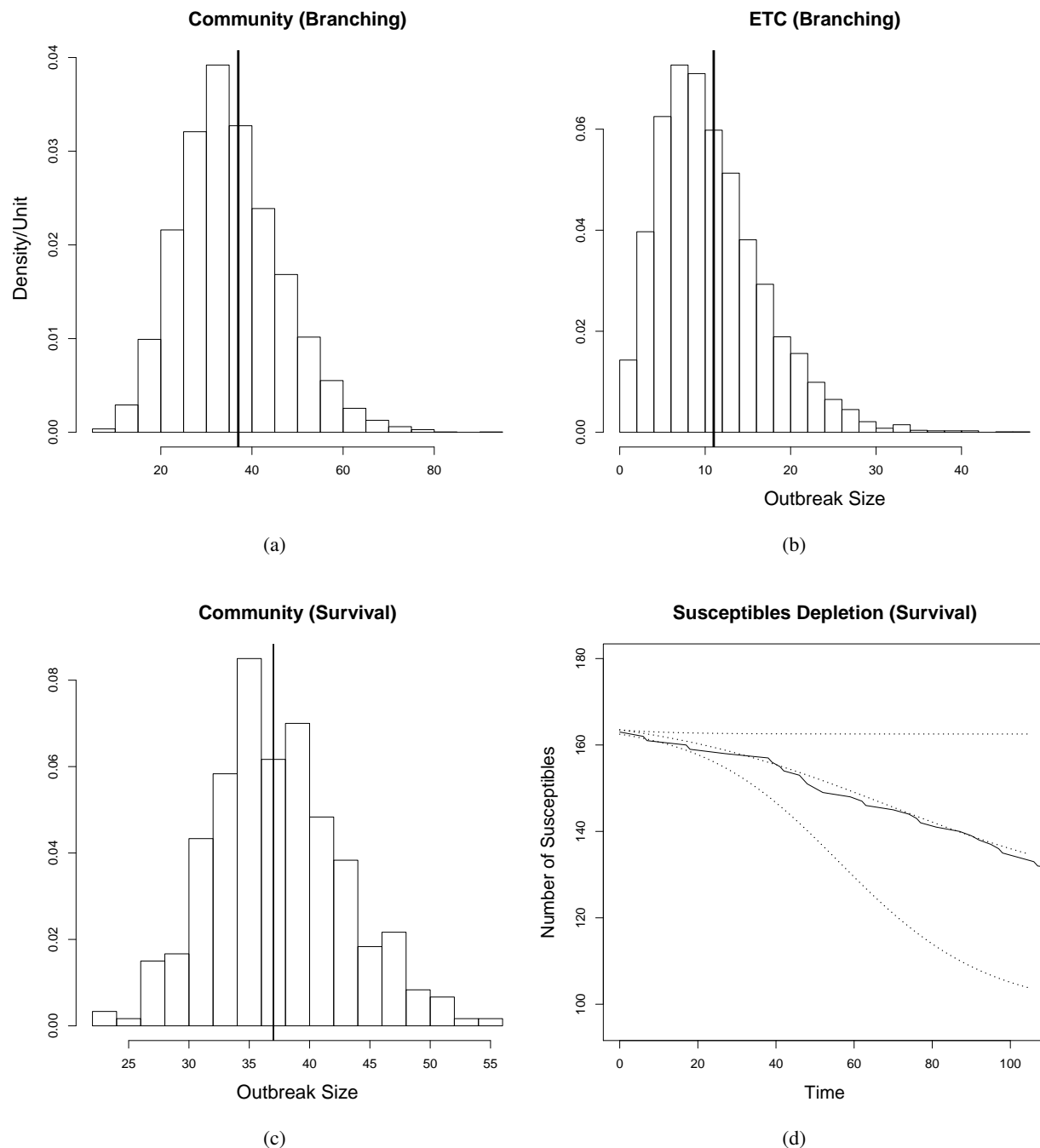


Fig. 2: **Model validation.** Top panels: branching process model predicted final size distributions and observed values (marked with vertical lines) for the (a) community and (b) ETC outbreaks. Bottom panels: (c) survival model predicted final size distribution and the observed value and (d) survival model predicted depletion of the susceptible population and the observed depletion. The curves in (d) are drawn conditionally on the estimated mean initial population size $n = 163$ with the lower and upper dotted lines representing model's 95% CI bounds.

- [3] P. S. Dodds and D. J. Watts, “Universal behavior in a generalized model of contagion,” *Physical Review Letters*, vol. 92, no. 21, pp. 218 701(1) – 218 201(4), 2004.
- [4] T. House and M. J. Keeling, “Insights from unifying modern approximations to infections on networks,” *Journal of The Royal Society Interface*, vol. 8, no. 54, pp. 67–73, 2011.
- [5] L. A. Meyers, B. Pourbohloul, M. E. Newman, D. M. Skowronski, and R. C. Brunham, “Network theory and SARS: predicting outbreak diversity,” *Journal of Theoretical Biology*, vol. 232, pp. 71–81, 2005.
- [6] M. E. Newman, “Spread of epidemic disease on networks,” *Physical Review E*, 2002.
- [7] M. Altmann, “The deterministic limit of infectious disease models with dynamic partners,” *Mathematical Biosciences*, vol. 150, no. 2, pp. 153–175, 1998.
- [8] K. Y. Leung, M. Kretzschmar, and O. Diekmann, “Dynamic concurrent partnership networks incorporating demography,” *Theoretical Population Biology*, vol. 82, pp. 229–239, 2012.
- [9] —, “SI infection on a dynamic partnership network: characterization of R_0 ,” *Journal of Mathematical Biology*, vol. 71, pp. 1–56, 2015.
- [10] M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, “Velocity and hierarchical spread of epidemic outbreaks in scale-free networks,” *Physical Review Letters*, vol. 92, no. 178701, 2004.
- [11] —, “Dynamical patterns of epidemic outbreaks in complex heterogeneous networks,” *Journal of Theoretical Biology*, vol. 235, pp. 275–288, 2005.
- [12] R. Pastor-Satorras and A. Vespignani, “Epidemic spreading in scale-free networks,” *Physical Review Letters*, vol. 86, no. 14, pp. 3200–3203, 2001.
- [13] J. Lindquist, J. Ma, P. van den Driessche, and F. H. Willeboordse, “Effective degree network disease models,” *Journal of Mathematical Biology*, vol. 62, no. 2, pp. 143–164, 2011.
- [14] J. Ma, P. van den Driessche, and F. H. Willeboordse, “Effective degree household network disease model,” *Journal of Mathematical Biology*, vol. 66, pp. 75–94, 2013.
- [15] E. Volz, “SIR dynamics in random networks with heterogeneous connectivity,” *Journal of Mathematical Biology*, vol. 56, no. 3, pp. 293–310, 2008.
- [16] J. C. Miller and E. M. Volz, “Incorporating disease and population structure into models of SIR disease in contact networks,” *PLoS One*, vol. 8, no. 8, p. e69162, 2013.
- [17] I. Z. Kiss, J. C. Miller, and P. L. Simon, *Mathematics of epidemics on networks: from exact to approximate models*. Springer, 2017, vol. 46.
- [18] W. R. KhudaBukhsh, B. Choi, E. Kenah, and G. A. Rempala, “Survival dynamical systems for the population-level analysis of epidemics,” *arXiv preprint arXiv:1901.00405*, 2019.
- [19] K. J. Olival and D. T. Hayman, “Filoviruses in bats: current knowledge and future directions,” *Viruses*, vol. 6, no. 4, pp. 1759–1788, 2014.
- [20] K. M. Johnson, J. V. Lange, P. A. Webb, and F. A. Murphy, “Isolation and partial characterisation of a new virus causing acute haemorrhagic fever in Zaire,” *Lancet*, vol. 309, no. 8011, pp. 569–571, 1977.
- [21] C. for Disease Control, P. (CDC) et al., “Outbreaks chronology: Ebola virus disease,” *CDC*, <http://www.cdc.gov/vhf/ebola/outbreaks/history/chronology.html>, 2014.
- [22] G. D. Maganga, J. Kapetshi, N. Berthet, B. Kabela Ilunga, F. Kabange, P. Mbala Kingebeni, V. Mondonge, J.-J. T. Muyembe, E. Bertherat, S. Briand et al., “Ebola virus disease in the Democratic Republic of Congo,” *New England Journal of Medicine*, vol. 371, no. 22, pp. 2083–2091, 2014.
- [23] C. E. M. Coltart, A. M. Johnson, and C. J. M. Whitty, “Role of healthcare workers in early epidemic spread of Ebola: policy implications of prophylactic compared to reactive vaccination polict in outbreak prevention and control,” *BMC Medicine*, vol. 13, p. 271, 2015.
- [24] H. Feldmann and T. W. Geisbert, “Ebola haemorrhagic fever,” *Lancet*, vol. 377, no. 9768, pp. 849–862, 2011.
- [25] P. Francesconi, Z. Yoti, S. Declich, P. A. Onek, M. Fabiani, J. Olango, R. Andraghetti, P. E. Rollin, C. Opira, D. Greco, and S. Salmaso, “Ebola hemorrhagic fever transmission and risk factors of contacts, Uganda,” *Emerging Infectious Diseases*, vol. 9, no. 11, pp. 1430–1437, 2003.
- [26] S. V. Scarpino, A. Iamarino, C. Wells, D. Yamin, M. Ndeffo-Mbah, N. S. Wenzel, S. J. Fox, T. Nyenswah, F. L. Altice, A. P. Galvani et al., “Epidemiological and viral genomic sequence analysis of the 2014 Ebola outbreak reveals clustered transmission,” *Clinical Infectious Diseases*, p. ciu1131, 2014.
- [27] S. F. Dowell, R. Mukunu, T. G. Ksiazek, A. S. Khan, P. E. Rollin, and C. Peters, “Transmission of Ebola hemorrhagic fever: a study of risk factors in family members, Kikwit, Democratic Republic of the Congo, 1995,” *Journal of Infectious Diseases*, vol. 179, no. Supplement 1, pp. S87–S91, 1999.
- [28] A. Matanock, M. A. Arwady, P. Ayscue, J. D. Forrester, B. Gaddis, J. C. Hunter, B. Monroe, S. K. Pillai, C. Reed, I. J. Schafer, M. Massaquoi, B. Dahn, and K. M. De Cock, “Ebola Virus Disease cases among health care workers not working in Ebola treatment units - Liberia, June-August, 2014,” *Morbidity and Mortality Weekly Report*, vol. 63, no. 46, pp. 1077–1081, 2014.
- [29] D. Yamin, S. Gertler, M. L. Ndeffo-Mbah, L. A. Skrip, M. Fallah, T. G. Nyenswah, F. L. Altice, and A. P. Galvani, “Effect of Ebola progression on transmission and control in Liberia,” *Annals of Internal Medicine*, vol. 162, no. 1, pp. 11–17, 2015.
- [30] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga et al., “Social contacts and mixing patterns relevant to the spread of infectious diseases,” *PLoS Med*, vol. 5, no. 3, p. e74, 2008.

- [31] A. S. Khan, F. K. Tshioko, D. L. Heymann, B. Le Guenno, P. Nabeth, B. Kerstiens, Y. Fleerackers, P. H. Kilmarx, G. R. Rodier, O. Nkuku, P. E. Rollin, A. Sanchez, R. Swanepoel, O. Tomori, S. T. Nichol, C. J. Peters, J.-J. Muyembe-Tamfum, and T. G. Ksiazek, "The reemergence of Ebola hemorrhagic fever, Democratic Republic of Congo, 1995," *Journal of Infectious Diseases*, vol. 179, no. S1, pp. S76–S86, 1999.
- [32] D. S. Chertow, C. Kleine, J. K. Edwards, R. Scaini, R. Giuliani, and A. Sprecher, "Ebola virus disease in West Africa—clinical manifestations and management," *New England Journal of Medicine*, vol. 371, no. 22, pp. 2054–2057, 2014.
- [33] A. M. Henao-Restrepo, I. M. Longini, M. Egger, N. E. Dean, W. J. Edmunds, A. Camacho, M. W. Carroll, M. Doumbia, B. Draguez, S. Duraffour et al., "Efficacy and effectiveness of an rVSV-vectored vaccine expressing Ebola surface glycoprotein: interim results from the Guinea ring vaccination cluster-randomised trial," *The Lancet*, vol. 386, no. 9996, pp. 857–866, 2015.
- [34] G. Chowell, N. W. Hengartner, C. Castillo-Chavez, P. W. Fenimore, and J. Hyman, "The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda," *Journal of Theoretical Biology*, vol. 229, no. 1, pp. 119–126, 2004.
- [35] P. E. Lekone and B. F. Finkenstädt, "Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study," *Biometrics*, vol. 62, no. 4, pp. 1170–1177, 2006.
- [36] J. Legrand, R. Grais, P. Boelle, A. Valleron, and A. Flahault, "Understanding the dynamics of Ebola epidemics," *Epidemiology and Infection*, vol. 135, no. 04, pp. 610–621, 2007.
- [37] J. M. Drake, I. Bakach, M. R. Just, S. M. O'Regan, M. Gambhir, and I. C.-H. Fung, "Transmission models of historical Ebola outbreaks," *Emerging Infectious Diseases*, vol. 21, no. 8, p. 1447, 2015.
- [38] J. S. Weitz and J. Dushoff, "Modeling post-death transmission of Ebola: Challenges for inference and opportunities for control," *Scientific Reports*, vol. 5, 2015.
- [39] C. M. Rivers, E. T. Lofgren, M. Marathe, S. Eubank, and B. L. Lewis, "Modeling the impact of interventions on an epidemic of Ebola in Sierra Leone and Liberia," *PLoS Currents*, vol. 6, 2014.
- [40] B. Tsanou, G. M. Moremedi, R. Kaondera-Shava, J. M. Lubuma, and N. Morris, "A simple mathematical model for Ebola in Africa," *Biomath Communications*, vol. 2, no. 1, 2015.
- [41] L. Valdez, H. H. A. Rêgo, H. Stanley, and L. Braunstein, "Predicting the extinction of Ebola spreading in Liberia due to mitigation strategies," *arXiv preprint arXiv:1502.01326*, 2015.
- [42] A. Pandey, K. E. Atkins, J. Medlock, N. Wenzel, J. P. Townsend, J. E. Childs, T. G. Nyenswah, M. L. Ndeffo-Mbah, and A. P. Galvani, "Strategies for containing Ebola in West Africa," *Science*, vol. 346, no. 6212, pp. 991–995, 2014.
- [43] M. F. Gomes, A. P. y Piontti, L. Rossi, D. Chao, I. Longini, M. E. Halloran, and A. Vespignani, "Assessing the international spreading risk associated with the 2014 West African Ebola outbreak," *PLoS Currents*, vol. 6, 2014.
- [44] S. Merler, M. Ajelli, L. Fumanelli, M. F. Gomes, A. P. y Piontti, L. Rossi, D. L. Chao, I. M. Longini, M. E. Halloran, and A. Vespignani, "Spatiotemporal spread of the 2014 outbreak of Ebola virus disease in Liberia and the effectiveness of non-pharmaceutical interventions: a computational modelling analysis," *The Lancet Infectious Diseases*, vol. 15, no. 2, pp. 204–211, 2015.
- [45] J. M. Drake, R. Kaul, L. W. Alexander, S. M. O'Regan, A. M. Kramer, J. T. Pulliam, M. J. Ferrari, and A. W. Park, "Ebola cases and health system demand in Liberia," *PLoS Biol*, vol. 13, no. 1, p. e1002056, 2015.
- [46] C. Wells, D. Yamin, M. L. Ndeffo-Mbah, N. Wenzel, S. G. Gaffney, J. P. Townsend, L. A. Meyers, M. Fallah, T. G. Nyenswah, F. L. Altice, K. E. Atkins, and A. P. Galvani, "Harnessing case isolation and ring vaccination to control Ebola," *PLoS Neglected Tropical Diseases*, 2015.
- [47] G. Webb, C. Browne, X. Huo, O. Seydi, M. Seydi, and P. Magal, "A model of the 2014 Ebola epidemic in West Africa with contact tracing," *PLoS Currents*, vol. 7, 2014.
- [48] D. Fisman, E. Khoo, and A. Tuite, "Early epidemic dynamics of the West African 2014 Ebola outbreak: estimates derived with a simple two-parameter model," *PLoS Currents*, vol. 6, 2014.
- [49] T. Kratz, P. Roddy, A. T. Oloma, B. Jeffs, D. P. Ciruelo, O. de la Rosa, and M. Borchert, "Ebola Virus Disease outbreak in Isiro, Democratic Republic of the Congo, 2012: signs and symptoms, management and outcomes," *PLoS ONE*, vol. 10, no. 6, p. e0129333, 2015.
- [50] M. G. Burch, K. A. Jacobsen, J. H. Tien, and G. A. Rempala, "Network-based analysis of a small ebola outbreak," *Math Biosci Eng.*, vol. 1, no. 14, pp. 67–77, 2017.
- [51] M. Newman, *Networks: an introduction*. Oxford University Press, 2010.
- [52] J. C. Miller, A. C. Slim, and E. M. Volz, "Edge-based compartmental modelling for infectious disease spread," *Journal of The Royal Society Interface*, vol. 9, no. 70, pp. 890–906, 2012. [Online]. Available: <http://rsif.royalsocietypublishing.org/content/9/70/890>
- [53] S. Chib and E. Greenberg, "Understanding the metropolis-hastings algorithm," *The American Statistician*, vol. 49, no. 4, pp. 327–335, 1995.
- [54] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*. CRC press Boca Raton, FL, 2014, vol. 2.