Qeios

Research Article

Poisson Network SIR Epidemic Model

Josephine K. Wairimu[1], Andrew Gothard[2], Grzegorz A. Rempala[2]

1. Department of Mathematics, University of Nairobi, Kenya; 2. Division of Biostatistics, The Ohio State University, Columbus, United States

We extend the classical Susceptible-Infected-Recovered (SIR) model to a network-based framework where the degree distribution of nodes follows a Poisson distribution. This extension incorporates an additional parameter representing the mean node degree, allowing for the inclusion of heterogeneity in contact patterns. Using this enhanced model, we analyze epidemic data from the 2018-20 Ebola outbreak in the Democratic Republic of the Congo, employing a survival approach combined with the Hamiltonian Monte Carlo method. Our results suggest that network-based models can more effectively capture the heterogeneity of epidemic dynamics compared to traditional compartmental models, without introducing unduly overcomplicated compartmental framework.

I. Introduction: SIR Compartmental Modeling

The study of the causes, transmission, and control of infectious diseases has a rich history, dating back to Grant's work in 1620[1]. In mathematical epidemiology, mathematical models and analytical techniques are employed to understand the spread and control of infectious diseases. These models enable researchers to predict disease dynamics, assess the impact of various interventions, and guide public health strategies.

One of the earliest examples of such modeling comes from Daniel Bernoulli[2], who in 1760 applied mathematical principles to study the trade-offs between the benefits and risks of variolation during a deadly smallpox outbreak in the American colonies. His approach classified individuals based on their epidemiological status, laying the groundwork for what are now known as compartmental models.

The first malaria compartmental model was developed by Sir Ronald Ross in early 1900's. He used human and vector compartments to demonstrate that malaria was transmitted by mosquitoes to humans[3]. In 1927, the classical Susceptible-Infected-Recovered (SIR) compartmental model was

introduced by Kermack and McKendrick to describe the transmission of infectious diseases[4][5][6]. For a comprehensive review of both classical and more recent SIR-type models, as well as related developments, see[1] and the references therein.

In the original Kermack-McKendrick model, individuals in a population are divided into three compartments: susceptibles (S) infected (I) and recovered (R), with the proportion of S, I, R type individuals at time t denoted by $S(t)$, $I(t)$, and $R(t)$ respectively. The system of ordinary differential equations (ODEs) governing the transfer of individuals between these compartments is given by

$$
\begin{aligned}
\dot{S} &= -bSI \\
\dot{I} &= bSI - cI \\
\dot{R} &= cI
\end{aligned}
\tag{1}
$$

where dotted quantities denote time derivatives and where for simplicity we suppressed the time argument in functions S, I, and R.

The above system is known as the SIR compartmental ODE model, the simplest example of a deterministic system describing the spread of a disease in a closed population. From the equations, we note the following.

- b represents the rate at which susceptible individuals become infected upon encountering infected individuals (the infection rate).

- c defines the rate at which infected individuals recover or succumb, thus leaving the infected compartment (the recovery rate).

- $\mathcal{R}_0 = b/c$ is a threshold parameter known as the *basic reproduction number* which represents the average growth rate of secondary cases in a completely susceptible population[1].

When $\mathcal{R}_0 > 1$, the epidemic will, at least initially, spread within the population as the number of infected individuals increases. In contrast, if $\mathcal{R}_0 \leq 1$, the number of infected individuals will decline and the epidemic will not spread.

If $0 < d \ll 1$ represents a small proportion of initially infected individuals, the initial conditions of the system are given by $(S(0), I(0), R(0)) = (1, d, 0)$. From the form of (1), it follows that the system satisfies the conservation law $S + I + R = 1 + d$. This and elementary manipulations of the first and the last equation in (1) give the following alternative representation of the classical SIR model

$$
\begin{aligned}
\dot{S} &= -bS(1 + d - S + \mathcal{R}_0^{-1} \log S) \\
\dot{I} &= -\dot{S} - cI \\
R &= 1 + d - I - S.
\end{aligned}
\tag{2}
$$

Note that the first equation in the system above involves only S. Since it describes the decline of susceptibles (and consequently the emergence of new cases), it is often referred to as the *epidemic curve equation* (see, for instance, [7]).

For the SIR model to effectively describe the disease dynamics in a given population of interest, it is typically necessary to estimate the parameters b, c, and d.

While compartmental ODE models are valuable tools in epidemiology, they have several notable limitations [8]. First, these models rely on the law of mass action, which assumes uniform social contact patterns across the population. This simplification neglects the inherent heterogeneity in individual behaviors and fails to account for changes in contact patterns as an epidemic unfolds, a crucial factor for diseases where person-to-person transmission is the primary mode of spread.

Additionally, as deterministic models, they overlook the dynamics of early outbreak phases, during which the number of infections is small and stochastic events play a significant role in shaping disease transmission. As disease dynamics grow more complex—such as when multiple types of infected or susceptible individuals are involved—these models become increasingly intricate, complicating interpretation and analysis.

Given these limitations, a different approach, often based on stochastic modeling, is frequently more effective [8][9]. While stochastic models are typically more challenging to formulate, they better accommodate heterogeneity and can simplify the overall structure by reducing the number of compartments. Furthermore, they inherently capture variability by incorporating random fluctuations, offering a more realistic representation of disease spread in diverse populations [10].

This paper explores the classical SIR model from a stochastic perspective, with a focus on its extension to the network-based SIR model, where transmission dynamics are governed by a static contact network. Specifically, we highlight the lesser-known relationships between the classical SIR model, its Poisson network counterpart, and the pairwise closure condition. Through a real data example, we demonstrate that while both models have similar computational complexity, the network-based model holds the advantage of explicitly accounting for and revealing the network's degree distribution.

The rest of the paper is organized as follows: In Section II, we provide background on network-based stochastic SIR models, focusing on configuration models, their pairwise representations, and dynamical survival approaches to model fitting. Following [7], we argue that in some cases the

classical SIR model serves as a good approximation of the Poisson SIR network model. In Section III, we address the statistical inference problem for stochastic network-based SIR models under a Poisson degree distribution and derive the relevant likelihood equations for parameter estimation. We then analyze the 2018-20 Ebola outbreak data in the Democratic Republic of the Congo, fitting the data to the derived network model and comparing the result to similar analysis based on the classical SIR approach presented in [11]. Finally, in Section IV, we discuss the results and present conclusions from our work.

II. Stochastic Network SIR Model

Stochastic network models address some of the challenges discussed earlier by providing a more realistic framework for capturing and studying the complexities of disease spread in structured populations. These models account for the variability in individual interactions and transmission dynamics. A stochastic model represents contact patterns during an epidemic as a graph, where nodes correspond to individuals, and edges denote potential transmission routes. The stochastic SIR epidemic process on a network of size n can be described as follows.

At the onset of the epidemic, m individuals are randomly selected as initially infectious. Each infectious individual remains in this state for a duration drawn from an exponential distribution with rate γ. During this period, the individual contacts their immediate neighbors according to a Poisson process with intensity β. If a contacted neighbor is susceptible, they become infectious immediately. Once the infectious period ends, the individual recovers and becomes immune to further infections. All infectious periods and Poisson processes are assumed to be independent (see, e.g., [12]).

In many applications, it is useful to track the spread of an epidemic while simultaneously constructing the underlying transmission network. This can be achieved by generating a random graph in tandem with modeling the epidemic's spread. The process, illustrated in Figure 1, begins by assigning to each node a number of unconnected half-edges based on the degree distribution p. Static connections are then formed by matching the half-edges of infectious nodes to other available half-edges in the network as part of the Poisson contact process with intensity β, as described earlier. If the connected node is susceptible, transmission occurs. Otherwise, the infectious individual attempts to connect its remaining half-edges to other available half-edges, repeating this process until all half-edges are matched or no further connections are possible. While this matching process may occasionally result

in self-loops or multiple edges between two nodes, such occurrences are negligible in the limit of large n.

The network SIR process as described above is seen to be a continuous time Markov chain [13] on a configuration model random graph $\mathcal{G}(n,p)$ of size n and the degree distribution $p = (p_k)_{k=0}^{\infty}$. Recall that the probability generating function (PGF) corresponding to p is defined as

$$\psi(x) = \sum_{k=0}^{\infty} p_k x^k.$$

The average degree is then defined in terms of PGF ψ as $\mu = \psi'(1)$. Similarly, the average excess degree (i.e., the average degree of a random neighbor of a node, minus one) is defined as $\mu_{ex} = \psi''(1)/\mu$ (see, e.g., [14][7]). The probability generating function is needed here as it tells us about the properties of a randomly chosen susceptible node. We use the PGF to find the probability that a randomly chosen node u, that is initially susceptible, in a infinite network, remains so up to time t.
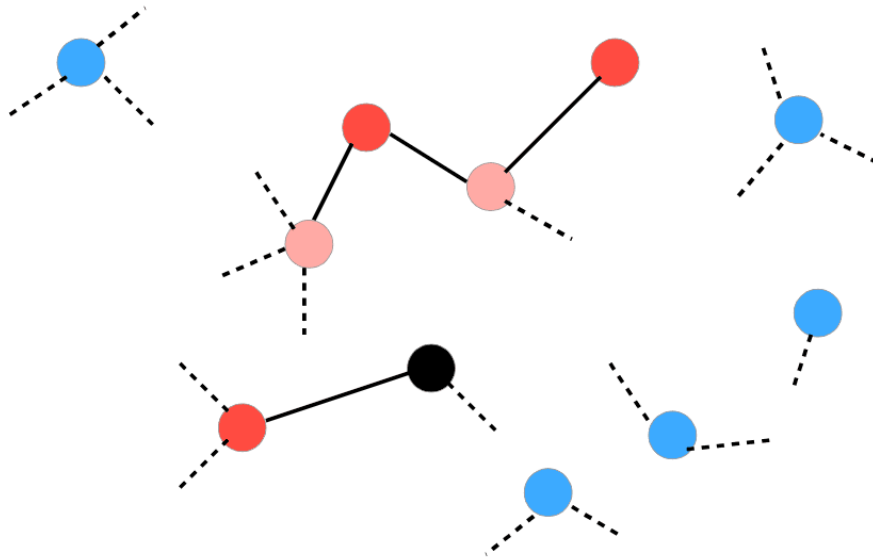

Figure 1. SIR Dynamics on Network. Blue nodes represent susceptible individuals, while red and pink ones represent the initially infected and secondarily infected individuals, respectively. The black node indicates a removed individual. Dashed half-edges connect uniformly at random to form solid edges.

A. Pairwise Model and Its Closure

The pairwise model provides a comprehensive way of describing the dynamics of an SIR epidemic on a configuration model graph. The pairwise model equations (see, for instance, [15]) are:

$$
\begin{aligned}
[\dot{S}] &= -\beta[SI] \\
[\dot{I}] &= \beta[SI] - \gamma[I] \\
[\dot{R}] &= \gamma[I] \\
[\dot{SI}] &= -\gamma[SI] + \beta([SSI] - [ISI]) - \beta[SI] \\
[\dot{SS}] &= -2\beta[SSI]
\end{aligned}
\tag{3}
$$

where $[A]$, $[AB]$, $[ABC]$ with $A, B, C \in \{S, I, R\}$ stand for the functions that at their integer values may be interpreted as counts of the number of singles, doubles and triples in the entire network with the given sequence of states when each group is counted in all possible ways. More formally, at integer values

$$
[ABC] = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} a_{ij} a_{jk} I_i(A) I_j(B) I_k(C)
$$

where $(a_{ij})_{i,j=1,2,\ldots,N}$ is the adjacency matrix of the network with entries either zero or one and $I_i(A)$, $I_i(B)$, and $I_i(C)$ are binary variables that equal one when the status of i-th individual is A, B, and C, respectively, and equal zero otherwise. The singles $[A]$ and doubles $[AB]$ are similarly defined.

The model described in (3) is not especially useful as stated, since clearly additional equations for triples dynamics are required, which, in turn, depend on dynamics of quadruples, and so on, leading to an escalating number of variables and equations. To manage this complexity and make the model tractable, a "closure" approach is often employed. In this approach, larger structures (e.g., triples) are approximated in terms of smaller ones (e.g., pairs). Indeed, The model described in (3) can be closed at the level of pairs by applying the following result recently presented in[15].

Theorem 1. *[Exact pairwise closure] Consider the SIR Markov process on the configuration graph $\mathcal{G}(n, p)$ as described above. Let $\kappa = \mu_{ex}/\mu$. The system (3) may be closed exactly by setting*

$$
[ASI] \simeq \kappa \frac{[AS][SI]}{[S]}
$$

for $A \in \{S, I\}$ iff the degree distribution p is binomial, Poisson or negative binomial. The closure is exact in the sense that the equality in (1) with both sides multiplied by n^{-1} holds asymptotically as $n \to \infty$. Furthermore, $\kappa = (k-1)/k < 1$, $\kappa = 1$, or $\kappa = (r+1)/r > 1$ if the degree distribution is binomial $\mathrm{BINOM}(k, q)$, Poisson $\mathrm{POI}(\lambda)$, or negative binomial $\mathrm{NB}(r, q)$, respectively.

We direct interested readers to the original paper for a formal statement of this result, along with its proof and additional details that are beyond the scope of this discussion.

Applying the closure relation in Theorem 1, the system (3) may be written as

$$
\begin{aligned}
[\dot{S}] &= -\beta[SI] \\
[\dot{I}] &= \beta[SI] - \gamma[I] \\
[\dot{R}] &= \gamma[I] \\
[\dot{SI}] &= -\gamma[SI] + \beta\kappa\frac{[SI][SS] - [SI]}{[S]} - \beta[SI] \\
[\dot{SS}] &= -2\beta\kappa\frac{[SI][SS]}{[S]}.
\end{aligned}
\tag{4}
$$

The above system may be considerably simplified, by removing last two equations. Indeed, by dividing the last equation by the first one and solving the resulting differential equation under the assumption that $[SS](0) = \mu[S](0)$, it follows that $[SS] = \mu[S]^{2\kappa}$. Substituting now this formula into the fourth equation and dividing again by the first one we obtain another differential equation for $[SI]$ as a function of $[S]$. That equation may be then solved explicitly, depending on the value of κ, yielding the reduced version of (4) which by Theorem 1 may be written in terms of the limiting proportions of $[S]/n$, $[I]/n$, and $[R]/n$ as $n \to \infty$ which we denote below by S, I, R. The final system has the following form:

$$
\begin{aligned}
-\dot{S} &= \begin{cases} \tilde{\beta}(S^\kappa - S^{2\kappa}) + \frac{\tilde{\gamma}}{1-\kappa}S(1 - S^{\kappa-1}) + \tilde{\rho}S^\kappa & \text{if } \kappa \neq 1 \\ \tilde{\beta}(S - S^2) + \tilde{\gamma}S\log S + \tilde{\rho}S & \text{if } \kappa = 1 \end{cases} \\
\dot{I} &= -\dot{S} - \gamma I \\
R &= 1 + \rho - S - I,
\end{aligned}
\tag{5}
$$

where the initial conditions are (per analogy with the classical SIR model (1) in Section I) $S(0) = 1, I(0) = \rho, R(0) = 0$ and $\tilde{\rho} = \beta\mu\rho, \tilde{\gamma} = \beta + \gamma$, and $\tilde{\beta} = \mu\beta$.

Note that the equations above may be interpreted as the mean field approximation to the scaled SIR Markov process evolving on the configuration model random graph $\mathcal{G}(n,p)$ as described in the beginning of this Section.

B. Poisson Network SIR

For the reminder of the paper we will consider only the special case of the Poisson network SIR model, namely when p distribution in $\mathcal{G}(n,p)$ is of the form $p_k = \exp(-\mu)\mu^k/k!$ for $k \geq 0$. This gives the PGF formula

$$
\Psi(x) = \exp\{\mu(x - 1)\}
$$

and consequently implies that $\mu = \Psi'(1) = \Psi''(1)/\mu = \mu_{ex}$, so that $\kappa = 1$ (see also Theorem 1). In this case the system (5) takes the form

$$
\begin{aligned}
\dot{S} &= -\tilde{\beta}S(1 + \rho - S + \tilde{\mathcal{R}}_0^{-1}\log S) \\
\dot{I} &= -\dot{S} - \gamma I \\
R &= 1 + \rho - S - I,
\end{aligned}
\tag{6}
$$

where $\tilde{\mathcal{R}}_0 = \tilde{\beta}/\tilde{\gamma}$ is a network analogue of the basic reproduction number discussed in Section I. In general, (6) is seen to be algebraically closely related to (2) as formalized in the following result (see also[7]).

Proposition 1. *Assume that we wish to approximate the dynamics of a Poisson network SIR epidemic given by (6) using the classical SIR equations (2). The two models' respective epidemic curve equations (and thus also respective S-curves) coincide iff $b = \tilde{\beta}$, $d = \rho$ and $c = \tilde{\gamma}$. In this case $\mathcal{R}_0 = \tilde{\mathcal{R}}_0$. Moreover, if $\mathcal{R}_0 < \mu$ than the true infection curve under the network model satisfies*

$$
\dot{I} = -\dot{S} - c(1 - \frac{\mathcal{R}_0}{\mu})I.
$$

$\square$

The result above indicates that the Poisson network SIR model will closely approximate the classical SIR model in terms of its Susceptible–Infected–Recovered (SIR) curves when the mean degree of the Poisson network (μ) is high. In this scenario, the factor $1 - \mathcal{R}_0/\mu$ approaches one. As the mean degree increases, the network model converges to the classical SIR model because the high connectivity reduces the impact of individual–specific interaction patterns. Consequently, the infection and recovery dynamics become more uniform, aligning closely with the simpler SIR framework. This is further illustrated in Figure 2 where the the I-curves of the Poisson network SIR model are plotted (in red) for several different values of μ. The lowest curve corresponds to $\mu = \infty$ which is the classical SIR model represented by ODEs (1) and (2).
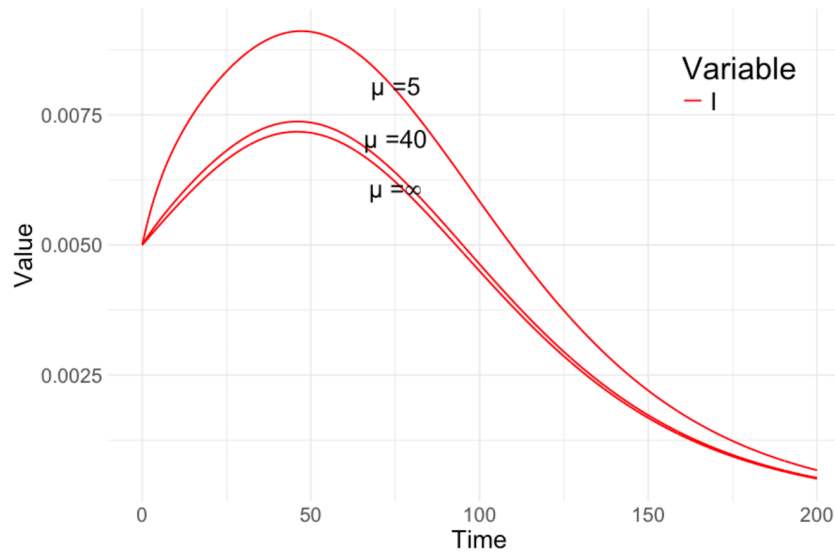
Figure 2. Approximating Network SIR Model. A simple example illustrating the result of Proposition 1. For large mean degree distribution μ (here at least 40), the network curves for infected are seen to get close to the one corresponding to the classical SIR model (lowest curve). For this particular example the values of the parameters $(\tilde{\beta}, \tilde{\gamma}, \rho)$ are taken from the first column of Table I in Section III below.

The system (6), for numerical convenience and to avoid evaluating terms of the form $\log S$ when S is small, is often reformulated into an alternative representation of the Poisson SIR equation. Specifically, by introducing the substitution

$$D = 1 + \rho - S + \tilde{\mathcal{R}}_0^{-1} \log S,$$

and differentiating, we can reformulate the first equation in terms of two others as shown in (7). These two equations resemble the classical SIR equations (1), with I replaced by D, and with the parameters matched as described in Proposition 1.

$$
\begin{aligned}
\dot{S} &= -\tilde{\beta} D S \\
\dot{D} &= -\dot{S} - \tilde{\gamma} D \\
\dot{I} &= -\dot{S} - \gamma I \\
R &= 1 + \rho - S - I,
\end{aligned}
\tag{7}
$$

with additional initial condition $D(0) = \rho$.

As noted in Chapter 2[16] for better numerical stability, it is sometimes convenient to rewrite the above set of equations in terms of log transformed quantities $S_* = \log S$, $I_* = \log I$, and $D_* = \log D$ as

$$\begin{aligned}
\dot{S}_* &= -\tilde{\beta} \exp(D_*) \\
\dot{D}_* &= \tilde{\beta} \exp(S_*) - \tilde{\gamma} \\
\dot{I}_* &= \tilde{\beta} \exp(S_* + D_* - I_*) - \gamma,
\end{aligned} \tag{8}$$

with R recovered via the conservation law as before.

C. Dynamical Survival Analysis

Although (6) is derived from the mean-field approximation of a Markov process evolving on an infinite random Poisson degree graph, it also serves as a basis for approximate inference regarding both the parameters of the underlying network and the SIR process. This method, which employs ODEs to characterize a stochastic process and is referred to as *dynamical survival analysis* or DSA in[9], was originally proposed in[17]. In that work, it was argued that, for a randomly selected node, the S function described in (6) can be interpreted as an improper survival function (see also[8]). Specifically, if T_I denotes the random time at which a randomly chosen, initially susceptible node in the network becomes infected, then in the limit of the large graph (that is, as the number of nodes $n \to \infty$) we have for any $t > 0$

$$P(T_I > t) = S(t).$$

Note that taking in (6) $\dot{S}(\infty) = 0$ (since at the end of the epidemic there is no further depletion of susceptibles, see e.g.,[8]) gives $S(\infty) = 1 - \tau$ where

$$\tau = \tilde{\mathcal{R}}_0^{-1} \log(1 - \tau) - \rho \tag{9}$$

is the final epidemic size or, equivalently, the limiting probability of a randomly chosen initially susceptible node being ever infected, i.e., $\tau = P(T_I < \infty)$.

More generally, if for $T > 0$ we define

$$\tau_T = 1 - S(T) \tag{10}$$

then the quantity

$$f_{\tau_T}(t) = -\dot{S}(t)/\tau_T \tag{11}$$

may be considered as a density function for the conditional random variable $T_I^* = T_I | T_I < T$. Notice that we may take here $T = \infty$ in which case $\tau_\infty = \tau$, the quantity defined in (9). In view of this, since

we may use the second equation in (6) to represent the variable I in terms of a convolution:

$$I(t) = \rho \exp(-\gamma t) + \int_0^t (-\dot{S}(u)) \exp\{-\gamma(t-u)\} du,$$

we see that the quantity

$$g_{\tau_T}(t) = \frac{\gamma}{\tau_T}\left(I(t) - \rho \exp(-\gamma t)\right) \tag{12}$$

is simply a density function of a random variable, say T_R, that is a sum of two independent random variables: T_I^* and U where U is an exponential variable with rate γ. It follows that the system (6) describes the densities of a pair of random variables (T_I^*, T_R) representing the history of epidemic of a randomly chosen network node in an infinite network. By construction, these variables are such that T_I^* is independent of the exponential variable $U = T_R - T_I^*$ which is interpreted as the infectious period.

Note that the vector of parameters describing the distribution of the pair (T_I^*, T_R) is

$$\theta = (\tilde{\beta}, \tilde{\gamma}, \gamma, \rho) \tag{13}$$

as all other relevant quantities, such as τ_T, μ, and $\tilde{\mathcal{R}}_0$, can be expressed in terms of θ or S, which is fully determined by θ. Indeed, recall that $\tilde{\mathcal{R}}_0 = \tilde{\beta}/\tilde{\gamma}$, $\tau_T = 1 - S(T)$ with τ_∞ defined by (9) and that for the average degree μ we have

$$\mu = \tilde{\mathcal{R}}_0 \left(1 - \frac{\gamma}{\tilde{\gamma}}\right)^{-1}. \tag{14}$$

In view of the above discussion, it is sometimes more convenient to consider an alternative parametrization

$$\theta^* = (\tilde{\beta}, \tilde{\gamma}, \mu, \rho), \tag{15}$$

which is the parametrization used in our numerical example below.

III. Statistical Inference

A. Ebola Dataset

We will illustrate the practical application of the network-based epidemic model and the approximation result in Proposition 1 by analyzing the data from the third and final wave of the 2018–2020 Ebola Virus Disease (EVD) outbreak in the Democratic Republic of the Congo (DRC). The

complete dataset from this outbreak is described and analyzed in detail in [11]; here, we focus specifically on a subset of the data, which includes information on the onset and hospitalization of 1,069 confirmed EVD patients recorded between May 27, 2019, and September 12, 2019. This subset is derived from the broader EVD database, which is comprehensively detailed in Section 2 of [11].

Since early August 2018, the DRC Ministry of Health, in collaboration with several international partners, has been working to support and strengthen the response to the Ebola Virus Disease (EVD) outbreak. This effort has been coordinated through the Emergency Operations Center in Goma, the capital of North Kivu Province in eastern DRC. Despite regional security challenges [18], response teams were deployed wherever possible to interview patients and suspected contacts. These interviews were conducted using a standardized case investigation form. Based on the collected information, cases were categorized as suspected, probable, or confirmed.

A suspected case (whether the patient survived or not) was defined as one with an acute onset of fever (over 100°F) and at least three Ebola-compatible clinical signs or symptoms, including headache, vomiting, anorexia, diarrhea, lethargy, stomach pain, muscle or joint aches, difficulty swallowing or breathing, hiccups, unexplained bleeding, or any sudden, unexplained death. If a patient met the suspected case definition but died, and no specimens were available for testing, the case was considered probable. A confirmed case of EVD was defined as a suspected case with at least one positive test for Ebola virus using reverse transcription polymerase chain reaction (RT-PCR) [19].

Patients suspected of having EVD were isolated and transported to an Ebola Treatment Center (ETC) for confirmatory testing and treatment [20]. Following the approach in [11], our analysis of the DRC dataset focused on the dates of symptom onset and removal, with "removal" defined as either death or recovery at home, or transfer to an ETC. It was assumed that, once patients were admitted to an ETC, the probability of further infection spread was minimal due to strict safety protocols. This assumption was later reinforced by the vaccination of healthcare personnel and family members who had been in contact with the suspected Ebola cases.

B. Statistical Model

The probabilistic DSA model used to derive (11) and (12) can now be applied in a similar manner to derive the *likelihood function* for inference of the parameters vector θ given in (13). The justification for this approach is discussed in [17] through the so-called Sellke construction. We omit the details here,

referring interested readers to the original source. Instead, our focus will be on the mechanics of the statistical inference process.

a. *Likelihood:* We assume that observations of new infections and recoveries are available up to some time horizon T such that $T \in [0, \infty]$. Recall the definition of (10) and note that τ_T is non-decreasing in T and that $\tau_\infty = \tau < 1$, the final epidemic size given by (9). As already indicated, the density (11) may be interpreted as a conditional density function of infection times T_I^* on $[0, T]$ for any $T \leq \infty$. Recalling the pair of variables (T_I^*, T_R) and the fact that the variable $W = T_R - T_I^*$ is exponentially distributed with rate parameter γ we obtain the following individual-level likelihood function for an initially susceptible individual i observed until time T with infection and recovery times t_i and r_i, respectively:

$$\mathcal{L}(\theta | t_i, r_i, T) = f_{\tau_T}(t_i) \gamma^{w_i} e^{-\gamma(r_i \wedge T - t_i)}. \tag{16}$$

Here, w_i is the event indicator, satisfying $w_i = 0$ if $r_i \wedge T = T$ and $w_i = 1$ otherwise. The likelihood for the set of n individuals in the population with complete records (16) is the product of the individual likelihoods, reflecting the assumption that the infection events are (approximately) independent in a large population.

b. *Likelihood with missing data:* As discussed in [11] in about 30% of the DRC Ebola cases the individual disease records were incomplete, missing either infection (t_i) or recovery (r_i) times. Fortunately, such missingness may be handled by the DSA likelihood without any need for data imputation. In case when only t_i is observed, (r_i is missing), the likelihood (16) reduces simply to (11)

$$\mathcal{L}(\theta | t_i, \circ, T) = f_{\tau_T}(t_i). \tag{17}$$

On the other hand, if only r_i is observed (t_i is missing), the likelihood is obtained from the convolution formula and (12)

$$\mathcal{L}(\theta | \circ, r_i, T) = g_{\tau_T}(r_i). \tag{18}$$

Similarly as above, the likelihood for incomplete data (22) is obtained by taking product of (17) and (18) over all individual incomplete histories (see also next subsection).

c. *Effective population size and outbreak size:* In many outbreak datasets, including the one from DRC considered here, only infection and recovery times are recorded. It is therefore often difficult to determine the size (N) of the susceptible population at risk of infection. This is known in the literature as the problem of estimating the *effective population size*[17]. Under the DSA model, this estimate may be obtained as

$$\hat{N} = \frac{k_T}{\tau_T}, \tag{19}$$

where k_T is the count of observed infected over the time horizon T and τ_T is given by (10). Similarly, one may also estimate the final epidemic count K_∞ of all already observed and future infections by

$$\hat{K}_\infty = \hat{N}\tau, \tag{20}$$

where $\tau = \lim_{T \to \infty} \tau_T$ is the final epidemic size.

C. Parameter Estimation

As discussed in the previous section, the epidemic history of the i-th individual can be represented either by the times of disease onset and removal (t_i, r_i) or by incomplete data, such as t_i or r_i alone, denoted as $(t_i, \circ)$ or $(\circ, r_i)$ ($\circ$ indicates a missing value). Among the available n_T individual recorded histories, we assume there are n complete records (t_i, r_i), n_1 incomplete records of type $(t_i, \circ)$, and n_2 incomplete records of type $(\circ, r_i)$.

The DSA likelihood function for the n complete data records is derived from (16) and takes the form:

$$\mathcal{L}_C(\theta | t_1, \ldots, t_n, r_1, \ldots, r_n, T) = \prod_{i=1}^{n} f_{\tau_T}(t_i) \gamma^{w_i} e^{-\gamma(r_i \wedge T - t_i)}, \tag{21}$$

where T is the time horizon, and w_i is the binary censoring indicator as defined in (16).

For the remaining $n_1 + n_2$ partially incomplete records, the DSA likelihood function is based on (11) and (12):

$$\mathcal{L}_I(\theta | t_1, \ldots, t_{n_1}, r_1, \ldots, r_{n_2}, T) = \prod_{i=1}^{n_1} f_{\tau_T}(t_i) \prod_{i=1}^{n_2} g_{\tau_T}(r_i), \tag{22}$$

where we assume $r_i < T$.

The overall likelihood for all n_T individual histories is obtained by multiplying (21) and (22). Note that the likelihood expressions depend on the parameter β only implicitly, through the values of the survival function $S(t)$ defined in (7).

The estimation of the parameter vector θ, as defined in (13), involves maximizing the product of the two likelihood functions (21) and (22). To enhance numerical stability, it is often advantageous to consider the formulation in terms of the logarithmic transformation (8), which involves maximizing the sum of the logarithms of the two likelihood functions instead. This approach seamlessly integrates into the Bayesian estimation framework, allowing for more comprehensive propagation of

uncertainty and the incorporation of external information into the statistical model. Consequently, this method yields parameter estimates that fully leverage all available information while accounting for uncertainty in a rigorous manner.

In our analysis of DRC data, the approximate posterior densities of θ were obtained using the Hamiltonian Monte Carlo (HMC) sampler[21], implemented in the open-source statistical software STAN[22]. This implementation was accessed via the R library Rstan[23]. Following[11], we assumed uniform (sometimes improper) prior distributions for the components of θ, specified as follows:

$$
\begin{aligned}
\tilde{\beta} &\in (0.1, \infty), \\
\tilde{\gamma} &\in (0, \beta), \\
\rho &\in (0, 0.01).
\end{aligned}
\tag{23}
$$

As in[11], the lower bound on $\tilde{\beta}$ was informed by empirical data, while the upper bound on γ was chosen to enforce the constraint $\tilde{\mathcal{R}}_0 > 1$.

The MCMC analysis was based on two chains of 3,000 iterations, with a burn-in period of 1,000 iterations. Convergence of the chains was assessed using Rubin's R statistic[23]. The analysis produced approximate samples from the posterior distribution of the parameter vector θ^* (see Figure 4 below).

D. Results

The general comparison of the parametric predictions of the DSA model with the empirical data in the DRC data set between May 27, 2019, and September 12, 2019 is given in Figure 3, where the scaled theoretical densities of the epidemic are plotted alongside the observed relative daily counts of infection (onset) and removal (hospitalization). As seen from the plots, the network model appears to capture well the empirically observed patterns of daily counts represented by the histogram bars. The 95% credible bounds around the model fit (shaded in red) are calculated based on the model parameter posterior distributions estimated via the MCMC algorithm with priors described in Section III-C. We note that, although the network DSA model curve fit appears quite similar to that presented in Section 2 of[11], the added benefit is the estimation of the distribution of an average degree interpreted as an average number of contacts per infected individual prior to removal. Since the mean degree is estimated as relatively high the Proposition 1 justifies also the use of classical SIR model from[11] for the DSA analysis.

The numerical results of the MCMC analysis are summarized in Table I with some of the posterior plots presented in Figure 4. In Table I, the posterior mean and corresponding credibility interval for each component of θ^* given by (15) are listed along with the estimated reproduction number ($\tilde{\mathcal{R}}_0$). Additionally, in the last two rows, the posterior mean estimate of the outbreak size (K_∞) as defined in (20) is reported along with its 95% credible bounds. The MCMC estimation scheme that produced the numerical values listed in the table was based on likelihood functions in equations (21) and (22) conditioned on the observation periods ($T = 108$). Since the posterior estimate of the ratio $\tilde{\mathcal{R}}_0/\mu \approx 1/40$ is relatively low, the result of Proposition 1 indicates that we should expect a reasonably good agreement between the classical mass-action SIR model (2) and the Poisson network model (6). Indeed, as seen from the entries of Table I, the point parameter values (posterior means) for the infection rate $\tilde{\beta}$, recovery rate ($\tilde{\gamma}$) and the initial prevalence of infection ρ all are numerically close for the classical and network models. In fact, since the respective credible intervals for $\tilde{\beta}$ and ($\tilde{\gamma}$) under network model contain the credible intervals under classical SIR, one would conclude that the posterior estimates distributions are not statistically different.

Param	Classical	Network
$\tilde{\beta}$	0.235 (0.218, 0.253)	0.229 (0.209, 0.259)
$\tilde{\gamma}$	0.214 (0.199, 0.230)	0.215 (0.197, 0.242)
ρ	0.0067 (0.0055, 0.0081)	0.0055 (0.0046, 0.0073)
μ	–	39.48 (7.93, 93.00)
$\tilde{\mathcal{R}}_0$	1.098 (1.061,1.135)	1.071 (1.034, 1.109)
$\hat{K}_\infty$	3481.41	3773.37
	(2877.416, 4155.878)	(3373.245, 4226.315)

Table 1. Parameter Estimates. Comparison of posterior estimates (means and 95% credibility bounds) from the classical SIR and network SIR models for the DRC Ebola dataset. The parameters between two models are matched according to Proposition 1.
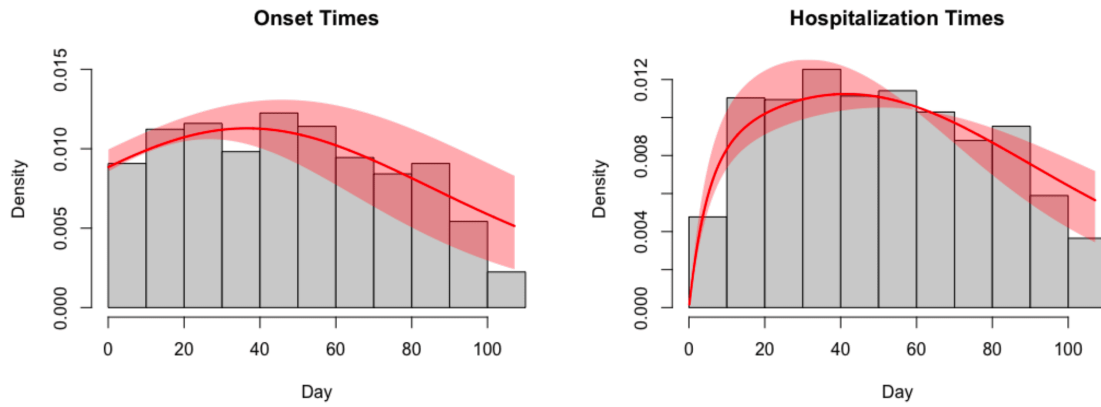
Figure 3. Model Fit. Histograms of onset and recovery (hospitalization) times overlaid with Poisson SIR model fit, computed at the means of the posterior parameter values (solid lines) with the corresponding 95% credible intervals (shaded regions).

Although the posterior means of the basic reproduction number and the final epidemic size are also close to the quantities obtained from the classical SIR model analysis, they posterior distributions are seen not to overlap and hence we conclude them to be different.

The final noteworthy result is that the SIR network model provides a better overall fit to the data, as indicated by the maximized likelihood function (not shown in the summary table). Additionally, the network analysis reveals that the posterior distribution of the average network degree (μ) is highly right-skewed. This finding aligns with the observation that a small subset of infectious individuals had a disproportionately high number of contacts before their hospitalization.

Interestingly, the mean contact degree is approximately 40, while the posterior mode is slightly below 25 (see Figure 4). For the range of parameters estimated from the data, the relatively high mean degree suggests that the two SIR models analyzed in Table I are numerically quite similar in terms of the respective numbers of infected individuals. This similarity is further illustrated by the plots in Figure 2 based on Proposition 1.
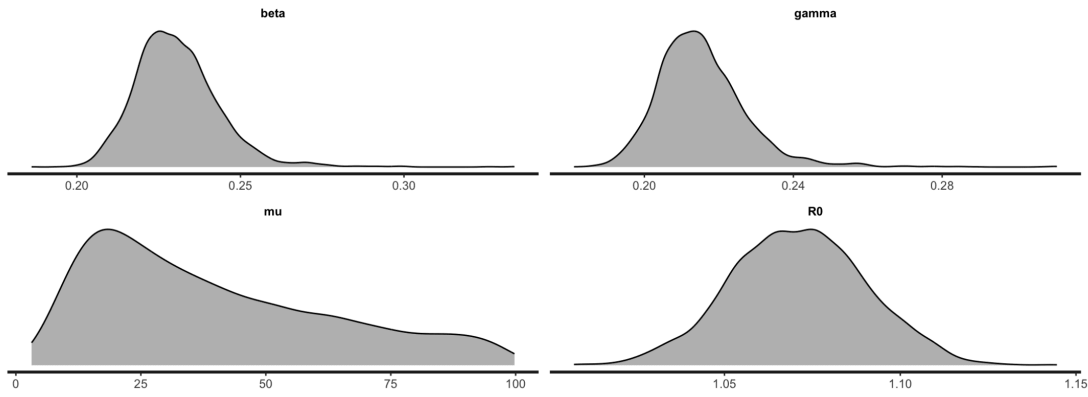
Figure 4. Posterior Parameter Densities. The posterior distributions of $\tilde{\beta}, \tilde{\gamma}, \mu$ and $\tilde{\mathcal{R}}_0$. The posterior distribution of average degree (μ) is seen to be right skewed but with mode below 25.

IV. Discussion

This paper presents a novel formulation of the classical SIR model, incorporating a dynamical survival analysis perspective into a configuration model based on a Poisson-distributed degree network. Among the broad class of configuration model networks, the Poisson network is particularly noteworthy due to its exact pair-level closure property. This characteristic makes it a representative model for many "appropriately regular" random networks (for instance, certain Bernoulli graphs) which often exhibit structural similarities to Poisson networks.

From a numerical perspective, the Poisson exact closure property prevents network complexity from scaling uncontrollably as the network size increases. This property enhances both analytical tractability and computational efficiency, offering a scalable and reliable approach for studying and managing disease dynamics. The proposed framework effectively models the contact patterns within a population that drive the spread of infectious diseases, providing deeper insights into outbreak mechanisms, disease persistence, and shifts in infection dynamics as interactions between infectious and susceptible individuals evolve.

The proposed model is broadly applicable to various domains, including social interactions, biological systems (e.g., neural or protein interactions), and technological networks (e.g., the spread of computer viruses or resilience of infrastructure systems). By transforming the SIR model using dynamical survival analysis within the edge-based configuration network framework, the resulting system of equations captures the intricate dynamics of network-based interactions. Despite the

complexity of these interactions, the equations remain mathematically tractable, often enabling precise predictions of disease trends (see, for instance, the discussions of related DSA-based approaches given in [8][24]).

The utility of the Poisson SIR network model is demonstrated through secondary analysis of the data from 2018–2020 Ebola outbreak in the Democratic Republic of the Congo. The model not only maintains a strong fit to empirical data but also reveals hidden structural features of the contact network underlying the disease's spread. Identifying such networks is crucial for effectively targeting at-risk populations—such as through vaccination campaigns—to prevent further transmission. This stochastic SIR framework thus provides a versatile tool for modeling infectious diseases and other dynamic processes beyond the scope of traditional SIR models.

Additionally, the study addresses an open challenge in approximating network-based SIR models with classical mass-action SIR models. Our empirical findings, encapsulated also in Proposition 1 reformulated after [7], demonstrate that such an approximation is achievable by establishing a correspondence between the parameters of the two frameworks. Furthermore, it highlights that in networks of moderate to high degree, the classical SIR model can serve as a reasonable approximation for the susceptible and infected curves and a possible trend prediction tool.

Overall, our results underscore the potential for bridging complex network-based dynamics with classical modeling approaches, offering both theoretical insights and practical tools for understanding and forecasting epidemic spread.

Statements and and Declarations

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contributions

GAR and JW developed the model, analyzed the data, and drafted the initial manuscript. AG contributed to data analysis, manuscript revisions, and the development of the necessary software scripts. All authors reviewed and approved the final version of the manuscript.

References

1. a, b, cBrauer F. Mathematical epidemiology: Past, present, and future. infect dis model. Infectious Disease Modeling. 2 (2): 113–127, February 2017.

2. ^Bernoulli D. Reflexions sur les avantages de l'inoculation. Mercure de Paris. 1760:173.

3. ^Ross R. The prevention of Malaria. Springer-Verlag; 1911.

4. ^Kermack WO, McKendrick AG (1927). "Contributions to the mathematical theory of epidemics ii". Proc R Soc London. 115: 700–721.

5. ^Kermack WO, McKendrick AG (1932). "Contributions to the mathematical theory of epidemics i. 1927." Proc R Soc London. 138: 55–83.

6. ^Kermack WO, McKendrick AG (1933). "Contributions to the mathematical theory of epidemics iii." Proc R Soc London. 141: 94-112.

7. a, b, c, d, eRempałla GA (2023). "Equivalence of mass action and poisson network sir epidemic models". Biomath. 12 (2311237).

8. a, b, c, d, eRempała GA, KhudaBukhsh WR. "Dynamical survival analysis for epidemic modeling", chapter Handbook of Visual, Experimental and Computational Mathematics: Bridges through Data, pages 1–17. Springer International Publishing, 2023.

9. a, bDi Lauro F, KhudaBukhsh WR, Kiss IZ, Kenah E, Max J, Rempała GA. "Dynamic survival analysis for non-markovian epidemic models". J. R. Soc. Interface. 19, June 2022.

10. ^Schwartz EJ, Choi B, Rempala GA (2015). "Estimating epidemic parameters: Application to h1n1 pandemic data". Mathematical Biosciences. 270: 198–203.

11. a, b, c, d, e, f, g, h, iVossler H, Akilimali P, Pan Y, KhudaBukhsh WR, Kenah W, Rempała GA (2022). "Analysis of individual-level data from 2018–2020 Ebola outbreak in Democratic Republic of the Congo". Scientific Reports. 12 (5534).

12. ^Gray A, Greenhalgh D, Hu L, Mao X (2011). "A stochastic differential equation SIS epidemic model". SIAM J. Appl. Math. **71**: 876–902.

13. ^Andersson T, Britton H. Stochastic epidemic models and their statistical analysis. Springer; 2000. Volume 151.

14. ^Newman MEJ (2002). "The spread of epidemic disease on networks". Physical Review E. **66** (016128).

15. a, b Kiss IZ, Kenah E, Rempala GA (2023). "Necessary and sufficient conditions for exact closures of epidemic equations on configuration model networks". Journal of Mathematical Biology. **87** (36).

16. ^Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. Princeton University Press; 2008.

17. a, b, c KhudaBukhsh WR, Choi B, Kenah E, Rempała GA (2020). "Survival dynamical systems: individual-level survival analysis from population-level epidemic models". Interface Focus. **10** (20190048).

18. ^Kraemer MUG, Pigott DM, Hill SC, Vanderslott S, Reiner RC, Stasse S, Brownstein JS, Gutierrez B, Dennig F, Hay SI, et al. Dynamics of conflict during the Ebola outbreak in the Democratic Republic of the Congo 2018–2019. BMC Medicine. **18**:1–10, 2020.

19. ^Maganga G, Kapetshi J, Berthet N, Kebela Ilunga B, Kabange F, Mbala Kingebeni P, Mondonge V, Muyembe JJ, Bertherat E, Briand S, et al. Ebola virus disease in the Democratic Republic of Congo. New England Journal of Medicine. **371**(22):2083–2091, 2014.

20. ^Oly Ilunga Kalenga, Moeti M, Sparrow A, Nguyen VK, Lucey D, Ghebreyesus TA (2019). "The ongoing Ebola epidemic in the Democratic Republic of Congo, 2018–2019". New England Journal of Medicine. **381** (4): 373–383.

21. ^Monnahan CC, Thorson JT, Branch TA (2017). "Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo". Methods in Ecology and Evolution. **8** (3): 339–348.

22. ^Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A (2017). "Stan: A probabilistic programming language". Journal of Statistical Software. **76** (1): 1–32.

23. a, b Annis J, Miller BJ, Palmeri TJ (2017). "Bayesian inference with Stan: A tutorial on adding custom distributions". Behavior Research Methods. **49** (3): 863–886.

24. ^Bastian CD, KhudaBukhsh WR, Klaus WM, Sahai SY, Root E, Weir MH, Kenah JH, Tien G, Rempała GA. "Projecting COVID-19 cases and hospital burden in Ohio". J Theor Biol. **561**(111404), 2023.

Declarations