IMPACT OF NON-MARKOVIAN RECOVERY ON NETWORK EPIDEMICS

GERGELY RÖST AND ZSOLT VIZI

Bolyai Institute,
University of Szeged,
Aradi vértanúk tere 1, Szeged 6720, Hungary
E-mail: rost@math.u-szeged.hu, zsvizi@math.u-szeged.hu

ISTVAN Z. KISS

School of Mathematical and Physical Sciences,
Department of Mathematics,
University of Sussex,
Falmer, Brighton BN1 9QH, UK
E-mail: i.z.kiss@sussex.ac.uk

We study how the distribution of infectious periods influences the dynamics of epidemics on networks. In our recently developed framework, we use pairwise models for network epidemics with non-Markovian recovery times. It is shown for typical families of distributions (such as gamma, uniform and lognormal) that higher variance in the recovery times generates lower reproduction numbers and different epidemic curves within each distribution family. We also show that knowing the expected value and the variance of the recovery times is not sufficient to determine the key characteristics of the epidemics such as initial growth rate, peak size, peak time and final epidemic size. For accurate predictions, more detailed information on the distribution of the infectious period is required, thus carefully estimating this distribution in the case of real epidemics has paramount public health importance.

1. Introduction

Networks (or graphs) offer a flexible framework to explicitly incorporate various heterogeneities in how individuals within a population interact [2,7,13,22,23]. This framework has led to a number of models where the strong assumptions of random mixing of the classical compartmental models can be relaxed. Because of the flexibility of the network approach, nodes can represent not only single individuals but also groups of individuals or locations. Similarly, links can represent contacts between individuals along

which diseases can spread, or interactions between groups such as flight routes between different locations.

In standard epidemiological models, the population is divided into classes depending on the status of individuals. In the simplest SIR (susceptible-infected-recovered) model, S, I and R denote the number (or proportion) of susceptible, infected and recovered individuals, respectively. The most important quantity associated with most epidemic models is the basic reproduction number (denoted by $\mathcal{R}_0$), which expresses the expected number of secondary infections generated by a 'typical' individual introduced into a susceptible population [8]. Naturally, if $\mathcal{R}_0 < 1$ the epidemic will die out, while if $\mathcal{R}_0 > 1$ the disease will spread. The other useful measure of epidemic severity is the final epidemic size, which is the total number of individuals who become infected during an epidemic, and often these two quantities can be related implicitly to capture factors such the network properties, and even properties of the transmission or recovery process (e.g. Poisson processes with exponentially distributed waiting time between events or general stochastic processes).

The transmission of epidemics on networks has been widely studied [23]. Most SIR models on networks assume that both the disease transmission and recovery process are Markovian. In other words, time to next infection along a link, where one node is susceptible and the other is infectious, is exponentially distributed and so is the time spent by an infectious node in the I state. Pairwise models have been very successful in capturing the average behaviour of a stochastic epidemics on networks [14]. These are formulated in terms of the expected values for the number of susceptible ($[S]$), infected ($[I]$) and recovered ($[R]$) nodes, which in turn depend on the expected values of (SS) pairs ($[SS]$) and (SI) pairs ($[SI]$) pairs. Introducing the following general notations

(1) $[X](t)$ is the expected number of nodes in state X at time t,
(2) $[XY](t)$ is the expected number of links which have a node in state X at one end and a node in state Y at the other, and
(3) $[XYZ](t)$ is the expected number of triplets in state $X - Y - Z$,

where, $X, Y, Z \in \{S, I, R\}$, and by accounting for all possible transitions,

the pairwise model is

$$
\begin{aligned}
[\dot{S}](t) &= -\tau[SI](t), \\
[\dot{I}](t) &= \tau[SI](t) - \gamma[I](t), \\
[\dot{SS}](t) &= -2\tau[SSI](t), \\
[\dot{SI}](t) &= \tau[SSI](t) - \tau[ISI](t) - \tau[SI](t) - \gamma[SI](t),
\end{aligned}
\tag{1}
$$

where τ is the per contact infection rate and γ is the rate of recovery. Furthermore, $[S] + [I] + [R] = N$ and above we only listed equations which are necessary to derive a full self-consistent system. It can be seen, that the system is not closed, because the equations for links contain triplets, thus we have to break the dependence on higher order terms. The closure approximation formula $[XSY] = \frac{n-1}{n}\frac{[XS][SY]}{[S]}$, where n is the average number of links per node and N is the number of nodes in the network, leads to the self-consistent system

$$
\begin{aligned}
[\dot{S}](t) &= -\tau[SI](t), \\
[\dot{I}](t) &= \tau[SI](t) - \gamma[I](t), \\
[\dot{SS}](t) &= -2\tau\frac{n-1}{n}\frac{[SS](t)[SI](t)}{[S](t)}, \\
[\dot{SI}](t) &= \tau\frac{n-1}{n}\left(\frac{[SS](t)[SI](t)}{[S](t)} - \frac{[SI](t)[SI](t)}{[S](t)}\right) - (\tau + \gamma)[SI](t).
\end{aligned}
\tag{2}
$$

Closing at the level of pairs with the approximation $[XY] = n[X]\frac{[Y]}{N}$, leads to the so called mean-field model

$$
\begin{aligned}
\dot{S}(t) &= -\tau\frac{n}{N}S(t)I(t), \\
\dot{I}(t) &= \tau\frac{n}{N}S(t)I(t) - \gamma I(t),
\end{aligned}
\tag{3}
$$

and the network structure is less important at this point. For this model the basic reproduction number is

$$
\mathcal{R}_0 = \frac{n}{N}\tau\mathbb{E}(\mathcal{I})S_0,
\tag{4}
$$

where, $\mathbb{E}(\mathcal{I}) = 1/\gamma$ is the expected infectious period. Various results for the Markovian case exist [14]. For the pairwise model, the final epidemic size is given by the following implicit relation

$$
\frac{s_\infty^{\frac{1}{n}} - 1}{\frac{1}{n-1}} = \frac{n-1}{N}\frac{\tau}{\tau + \gamma}[S]_0\left(s_\infty^{\frac{n-1}{n}} - 1\right),
\tag{5}
$$

4

where $[S]_0$ is the number of susceptible individuals at time $t = 0$ and $s_\infty = [S]_\infty/[S]_0$, where $[S](\infty) = [S]_\infty$. The assumption of Markovianity in both infection and recovery is a strong simplifying assumption, as especially in the context of epidemiology, the period of infectiousness has paramount importance [15,18], and often this is approximated from the empirical distribution of observed infectious periods of various diseases by log-normal and gamma (smallpox [9,20]), fixed-length (measles [1]) or Weibull distributions (ebola [4]). Newman [21], with his results later made more precise by Kenah and Robins [16], gives a relation between the average transmission rate and final epidemic size, using a bond percolation approach, and his result is applicable for more general infection and recovery processes. Recently, however there is renewed interest in modelling non-Markovian processes, such as epidemics on networks [3,5,6,12,16,17,19,24,25], and random walks [11]. Kiss et al. [17] considered an SIR epidemic with deterministic infectious period (i.e. the recovery time is constant) and derived the mean-field and pairwise models for this case. The mean-filed model is given by, see [17],

$$S'(t) = -\tau \frac{n}{N} S(t) I(t),$$
$$I'(t) = \tau \frac{n}{N} S(t) I(t) - \tau \frac{n}{N} S(t - \sigma) I(t - \sigma). \tag{6}$$

The pairwise model for fixed infectious period was derived [17] as

$$[\dot{S}](t) = -\tau[SI](t),$$
$$[\dot{SS}](t) = -2\tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)},$$
$$[\dot{I}](t) = \tau[SI](t) - \tau[SI](t - \sigma),$$
$$[\dot{SI}](t) = \tau \frac{n-1}{n} \frac{[SS](t)[SI](t)}{[S](t)} - \tau \frac{n-1}{n} \frac{[SI](t)[SI](t)}{[S](t)} - \tau[SI](t)$$
$$\qquad -\tau \frac{n-1}{n} \frac{[SS](t-\sigma)[SI](t-\sigma)}{[S](t-\sigma)} e^{-\int_{t-\sigma}^{t} \tau \frac{n-1}{n} \frac{[SI](u)}{[S](u)} + \tau du}. \tag{7}$$

We note that both models are now delay differential equations rather than ordinary differential equations (ODEs), as is the case for Markovian epidemics. In the same paper Kiss et al. have shown that the for the final size the following relation

$$\frac{s_\infty^{\frac{1}{n}} - 1}{\frac{1}{n-1}} = \frac{n-1}{N} \left(1 - e^{-\tau\sigma}\right) [S]_0 \left(s_\infty^{\frac{n-1}{n}} - 1\right) \tag{8}$$

holds, where a newly introduced basic reproduction-like number is defined as $\mathcal{R}_0^p = \frac{n-1}{N} \left(1 - e^{-\tau\sigma}\right) [S]_0$. Furthermore, in [17] it has been presented that

for arbitrary infectious periods, the basic reproduction number associated to the pairwise model is

$$\mathcal{R}_0^p = \frac{n-1}{N} \left(1 - \mathcal{L}[f_{\mathcal{I}}](\tau)\right) [S]_0, \tag{9}$$

where $\mathcal{L}[\cdot]$ is the Laplace transform and $f_{\mathcal{I}}$ is the probability density function of the recovery process. With this new definition, numerical tests (and in special cases analytical results) have confirmed, see [17], that the following implicit relation for the final epidemic size holds

$$\frac{s_\infty^{\frac{1}{n}} - 1}{\frac{1}{n-1}} = \mathcal{R}_0^p \left(s_\infty^{\frac{n-1}{n}} - 1\right)$$

$$= \frac{n-1}{N} \left(1 - \mathcal{L}[f_{\mathcal{I}}](\tau)\right) [S]_0 \left(s_\infty^{\frac{n-1}{n}} - 1\right), \tag{10}$$

and this can be applied for any general recovery time distributions.

Notice, while $\mathcal{R}_0$ depends on the expected value only, see (4), the pairwise reproduction number (9), uses the complete density function, thus the average length of infectious period does not determine exactly the reproduction number. This remark implies, that for an epidemic we have to know as precisely as possible the shape of the distribution. In the sequel we denote by $\mathcal{I}$ the random variable that describes the recovery time, which is, in the SIR-setting, the same as the distribution of the infectious period.

In this paper we exploit the generalised basic reproduction number (9) and the general implicit relation for the final epidemic size (10) to determine how these depend on the precise shape of recovery time distribution. We give some analytic results, as well as results based on simulations to shed light on the time evolution of the epidemics. The paper is structured as follows. Sections 2, 3, and 4 concern with recovery times of gamma, uniform and lognormal distributions, respectively. The dependence of the reproduction number, as well as the shape of the epidemic curve, on the distribution parameters is analysed in details in each case. In Section 5, by comparing epidemics generated by different types of the recovery time distributions, we illustrate that estimating the expected value and the variance of the recovery time is not sufficient to give a realistic description of the epidemics and more detailed knowledge of the shape of the empirical distribution is needed to give a reasonable approximation of a real epidemics.

2. Gamma distributed recovery time

The gamma distribution is one of the most commonly used distributions in the epidemiology literature to approximate empirically observed latent
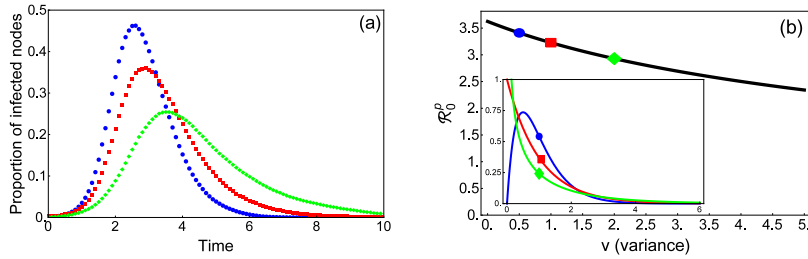
6

Figure 1. **(a)** Epidemic curves as averages of explicit stochastic simulations for non-Markovian epidemics, where the transmission rate is $\tau = 0.3$ and the initial number of susceptibles is $S_0 = 999$ on a homogeneous network with $N = 1000$ nodes and degree $n = 15$. The circles/squares/diamonds correspond to simulations for gamma distributed recovery time with parameters $(a, b) = (2, 0.5)/(1, 1)/(0.5, 2)$, respectively. **(b)** The solid curve shows the reproduction number $\mathcal{R}_0^p$ as a function of variance v for fixed $m = 1$, and the circle/square/diamond represent the cases simulated in Fig. (a). In the inset figure, the shapes of the three corresponding probability density functions are presented.

periods and infectious periods, because of its flexibility and the possibility of incorporating it into ordinary differential equation models by the method of stages (also called linear chain) [18] . For example, it has been fitted to the incubation period and infectious period of smallpox [9], bluetongue disease [10] and so on.

The usual notation of gamma distribution is $\mathrm{Gamma}(a, b)$, where a is called the shape parameter and b is called the scale parameter. The probability density function and its Laplace transform are

$$f_{\mathcal{I}}(x) = \frac{x^{a-1}e^{-\frac{x}{b}}}{\Gamma(a)b^a}, \quad \mathcal{L}[f_{\mathcal{I}}](s) = \left(\frac{1}{1+sb}\right)^a,$$

where $\Gamma(a)$ is the gamma function evaluated at a. If the infectious period $\mathcal{I}$ is gamma distributed with shape parameter a and scale parameter b, that is $\mathcal{I} \sim \mathrm{Gamma}(a, b)$, then the expected value is $m := \mathbb{E}(\mathcal{I}) = ab$, and the variance is $v := \mathrm{Var}(\mathcal{I}) = ab^2$, and for simplicity later we shall use the notation m and v to denote the mean (expected value) and the variance of distributions.

Proposition 2.1. *Consider two random variables $\mathcal{I}_1 \sim \mathrm{Gamma}(a_1, b_1)$ and $\mathcal{I}_2 \sim \mathrm{Gamma}(a_2, b_2)$ such that $\mathbb{E}(\mathcal{I}_1) = \mathbb{E}(\mathcal{I}_2)$ and $\mathrm{Var}(\mathcal{I}_1) \leq \mathrm{Var}(\mathcal{I}_2)$. If $\mathcal{I}_1$ and $\mathcal{I}_2$ represent the recovery time distribution, then for the corresponding reproduction numbers the relation $\mathcal{R}_{0,\mathcal{I}_1}^p \geq \mathcal{R}_{0,\mathcal{I}_2}^p$ holds (i.e. for gamma distributions with a given mean, the pairwise reproduction number is monotonically decreasing in variance).*
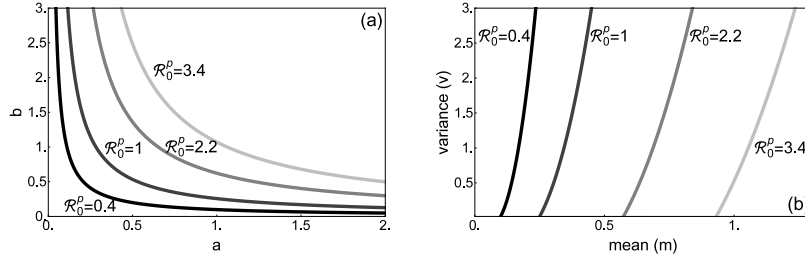
Figure 2.   Contour lines of $\mathcal{R}_0^p$ as a two-variable function of the parameters of the gamma distribution. The transmission rate is $\tau = 0.3$, the network has $N = 1000$ nodes and degree $n = 15$ and initial number of susceptibles is $S_0 = 999$. In **(a)**, the contour lines are given as the function of the shape and scale parameter, while in **(b)** they are depicted as the function of the mean and the variance of the gamma distribution.

Proof. Fix m as the same mean of $\mathcal{I}_1$ and $\mathcal{I}_2$, then the scale parameter can be expressed as $b = \frac{m}{a}$. Using that $v = \text{Var}(\mathcal{I}) = ab^2$ and

$$\mathcal{L}[f_{\mathcal{I}}](\tau) = \left(\frac{1}{1+\tau b}\right)^a,$$

we can express the parameters in terms of the mean m and variance v, and thus the Laplace transform can be written as

$$\lambda_m(\tau; v) := \mathcal{L}[f_{\mathcal{I}}](\tau) = \left(\frac{1}{1+\frac{\tau}{m}v}\right)^{\frac{m^2}{v}},$$

where the notation $\lambda_m(\tau; v)$ for the Laplace transform is meant to emphasize that the Laplace transform evaluated at τ for a fixed m is a function of v. For arbitrary $0 < x < y < 1$ and $0 < a < b$, the inequalities

$$x^{\frac{m^2}{a}} < x^{\frac{m^2}{b}}, \quad x^{\frac{m^2}{b}} < y^{\frac{m^2}{b}}$$

hold. For $v_1 \leq v_2$ and for $\tau > 0$ we have $\frac{1}{1+\frac{\tau}{m}v_2} \leq \frac{1}{1+\frac{\tau}{m}v_1}$, and using the relations above we obtain

$$\left(\frac{1}{1+\frac{\tau}{m}v_1}\right)^{\frac{m^2}{v_1}} \leq \left(\frac{1}{1+\frac{\tau}{m}v_1}\right)^{\frac{m^2}{v_2}} \leq \left(\frac{1}{1+\frac{\tau}{m}v_2}\right)^{\frac{m^2}{v_2}},$$

which means that $\lambda_m(\tau; v)$ is monotone increasing in v. Therefore, the pairwise reproduction number

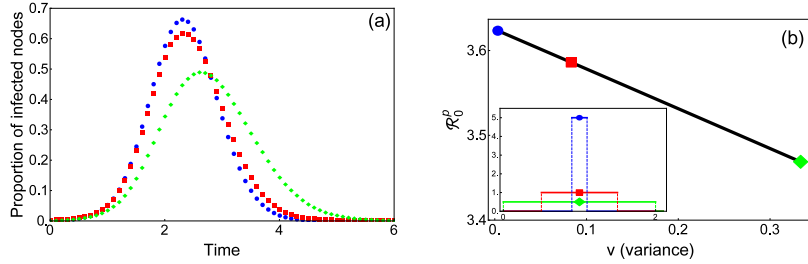$$R_{0,\mathcal{I}}^p = \frac{n-1}{N}[S]_0(1 - \lambda_m(\tau; v))$$

8

Figure 3. **(a)** Epidemic curves as averages of explicit stochastic simulations for non-Markovian epidemics, where transmission rate is $\tau = 0.3$ and initial number of susceptibles is $S_0 = 999$, on homogeneous network with $N = 1000$ nodes and degree $n = 15$. The circles/squares/diamonds correspond to simulations for uniformly distributed recovery time with parameters $(a, b) = (0.9, 1.1)/(0.5, 1.5)/(0, 2)$, respectively. **(b)** The solid curve shows the reproduction number $\mathcal{R}_0^p$ as the function of variance v for fixed $m = 1$, and the circle/square/diamond represent the cases simulated in Fig. (a). In the inset figure, the three uniform density functions are depicted.

is monotone decreasing in v. □

The monotonicity of the reproduction number in the variance is depicted in Fig. 1(b). For a fixed mean but different variances of the gamma distribution, we can observe different epidemic curves in Fig. 1(a), and correspondingly different reproduction numbers (see Fig. 1(b)). The dependence of $R_{0,\mathcal{I}}^p$ on the distribution parameters is detailed in Fig. 2.

3. Uniformly distributed recovery time

Since its simplicity allows us to make explicit calculations, in this Section we outline how the reproduction number and the disease dynamics behave when the recovery time follows uniform distribution. Uniformly distributed incubation and infectious periods have been used in the modeling of avian influenza [26]. Let Uniform(a, b) denote a uniform distribution corresponding to the interval $[a, b]$, where $a \geq 0, b > a$. If $\mathcal{I} \sim$ Uniform(a, b), then the expected value is $m = \mathbb{E}(\mathcal{I}) = \frac{a+b}{2}$, and the variance is $v = \text{Var}(\mathcal{I}) = \frac{(b-a)^2}{12}$. The probability density function and its Laplace transform are given as

$$f_{\mathcal{I}}(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}, \quad \mathcal{L}[f_{\mathcal{I}}](s) = \frac{e^{-sa} - e^{-sb}}{s(b-a)}.$$
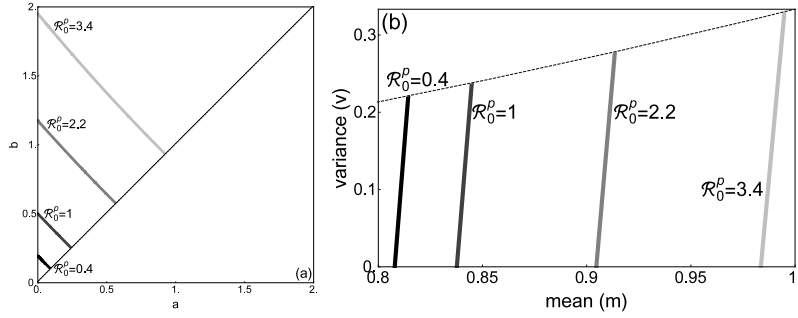
Figure 4.   **(a)** Contour lines of $\mathcal{R}_0^p$ as a two-variable function of the parameters of the uniform distribution. **(b)** Contour lines of $\mathcal{R}_0^p$ as a two-variable function of m and v for uniform distribution. For both (a) and (b), the transmission rate is $\tau = 0.3$, the network has $N = 1000$ nodes and degree $n = 15$ with an initial number of susceptibles $S_0 = 999$.

Proposition 3.1. *Consider two random variables $\mathcal{I}_1 \sim \mathrm{Uniform}(a_1, b_1)$ and $\mathcal{I}_2 \sim \mathrm{Uniform}(a_2, b_2)$ such that $\mathbb{E}(\mathcal{I}_1) = \mathbb{E}(\mathcal{I}_2)$ and $\mathrm{Var}(\mathcal{I}_1) \leq \mathrm{Var}(\mathcal{I}_2)$. If $\mathcal{I}_1$ and $\mathcal{I}_2$ represent the recovery time distribution, then for the corresponding reproduction numbers the relation $\mathcal{R}_{0,\mathcal{I}_1}^p \geq \mathcal{R}_{0,\mathcal{I}_2}^p$ holds (i.e. for uniform distributions with a given mean, the pairwise reproduction number is monotonically decreasing in variance).*

Proof.   Fixing the mean m, the right endpoint of the interval is $b = 2m - a$. Using that $v = \mathrm{Var}(\mathcal{I}) = \frac{(b-a)^2}{12}$ and $\mathcal{L}[f_{\mathcal{I}}](\tau) = \frac{e^{-\tau a} - e^{-\tau b}}{\tau(b-a)}$, by simple algebra we can express the parameters by m and the variance v, and consequently the Laplace transform can be written as the function of v:

$$\lambda_m(\tau; v) = \frac{e^{-\tau m}}{2m} \frac{e^{\tau\sqrt{3v}} - e^{-\tau\sqrt{3v}}}{\tau\sqrt{3v}},$$

where $v \in \left(0, \frac{m^2}{3}\right]$ from our assumptions on a and b. Expanding the exponentials in $\lambda_m(\tau; v)$ into Taylor series, we can notice that the negative terms cancel out, and we obtain

$$
\begin{aligned}
\lambda_m(\tau; v) &= \frac{e^{-\tau m}}{2m} \frac{\left(1 + (\tau\sqrt{3v}) + \dots\right) - \left(1 - (\tau\sqrt{3v}) + \dots\right)}{\tau\sqrt{3v}} \\
&= \frac{e^{-\tau m}}{2m} \frac{2\tau\sqrt{3v} + 2\frac{(\tau\sqrt{3v})^3}{3!} + \dots}{\tau\sqrt{3v}} \\
&= \frac{e^{-\tau m}}{m} \sum_{n=0}^{\infty} \frac{(\tau\sqrt{3v})^{2n}}{(2n+1)!},
\end{aligned}
\tag{11}
$$

10

which is monotone increasing in v for $m > 0$ and $\tau > 0$. Therefore, the pairwise reproduction number $R_{0,\mathcal{I}}^p = \frac{n-1}{N}[S]_0(1 - \lambda_m(\tau; v))$ is monotone decreasing in v. $\square$

The monotonicity of the reproduction number in the variance is depicted in Fig. 3(b). Similarly to the gamma distribution, for a fixed mean but different variances of the uniform distribution we can observe different epidemic curves in Fig. 3(a), and correspondingly different reproduction numbers (see Fig. 3(b)). The dependence of $R_{0,\mathcal{I}}^p$ on the distribution parameters is detailed in Fig. 4.
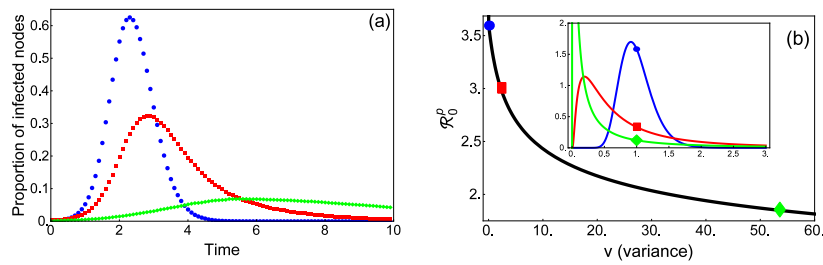
4. Lognormal distributed recovery time

Figure 5. **(a)** Epidemic curves as averages of explicit stochastic simulations for non-Markovian epidemics, where transmission rate is $\tau = 0.3$ and initial number of susceptibles is $S_0 = 999$, on homogeneous network with $N = 1000$ nodes and degree $n = 15$. The circles/squares/diamonds correspond to simulations for lognormally distributed recovery time with parameters $(\mu, \sigma) = (-0.03125, 0.25)/(-\frac{\ln(3)}{2}, \sqrt{\ln 3})/(0, 2)$, respectively. **(b)** The solid curve shows the reproduction number $\mathcal{R}_0^p$ as the function of variance v for fixed $m = 1$, and the circle/square/diamond represent the cases simulated in Fig. (a). Inset figure shows the shape of these three distributions.

The lognormal distribution is also widely used in epidemiology. They have been fitted, among others, to the incubation and infectious periods of smallpox [20]. Let $\ln\mathcal{N}(\mu, \sigma^2)$ denote a lognormal distribution, i.e. its logarithm is a normal distribution with expected value μ and variance σ. Then for the lognormal distribution $m = \mathbb{E}(\mathcal{I}) = e^{\mu + \frac{\sigma^2}{2}}$, $v = \text{Var}(\mathcal{I}) = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$, and the probability density function is, for $x > 0$,

$$f(x) = \frac{1}{x\sqrt{2\pi}\sigma}e^{-\frac{(-\mu + \ln(x))^2}{2\sigma^2}}.$$

Unfortunately a closed form formula does not exist for its Laplace transform, thus we can not repeat the analysis of the previous two sections.

However, we can still investigate numerically the impact of m and v on the reproduction number and the time course of the epidemic. The density function can again be expressed in terms of m and v by the formula

$$f_m(x;v) = \frac{1}{x\sqrt{2\pi}\sqrt{\ln\left(\frac{v}{m^2}+1\right)}}e^{-\frac{\left(\ln(x)-\ln(m)+\frac{1}{2}\ln\left(\frac{v}{m^2}+1\right)\right)^2}{2\ln\left(\frac{v}{m^2}+1\right)}}, \text{ for } x > 0. \quad (12)$$

By straightforward calculation, we can find $\mu = \ln(m) - \frac{1}{2}\ln\left(1+\frac{v}{m^2}\right)$ and $\sigma^2 = \ln\left(1+\frac{v}{m^2}\right)$, and then the formula above can be derived. Using this formula for the density, we can plot the numerically determined pairwise reproduction number as a function of the variance for any given m, see Fig. 5(b). The epidemic curves corresponding to these distributions can be seen in Fig. 5(a), and the dependence of $\mathcal{R}_0^p$ on the distribution parameters is detailed in Fig. 6.
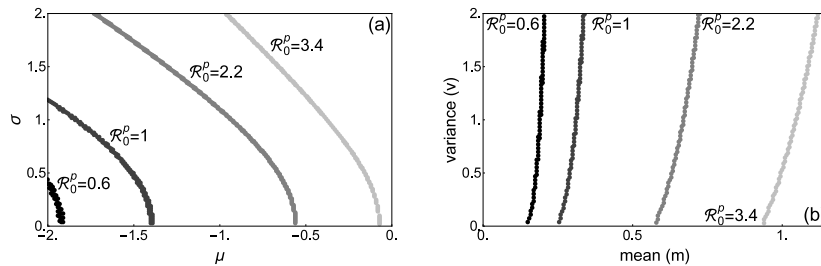

Figure 6.   **(a)** Contour lines of $\mathcal{R}_0^p$ as a two-variable function of the parameters of the lognormal distribution. **(b)** Contour lines of $\mathcal{R}_0^p$ as a two-variable function of the mean and variance for lognormal distribution. For both (a) and (b), the transmission rate is $\tau = 0.3$, the network has $N = 1000$ nodes and degree $n = 15$ with an initial number of susceptibles is $S_0 = 999$.

5. Discussion

We used our recently developed non-Markovian pairwise model [17] for network epidemics to investigate the impact of the shape of the distribution of the recovery times on SIR epidemics. In particular, we utilized our formula for the reproduction number which includes the Laplace transform of the probability density distribution of the infectious period[17]. We chose three families of common distributions (gamma, uniform and lognormal), and in each case we showed that (for the first two analytically, for the third one only numerically) if the mean recovery time is given, then smaller variance
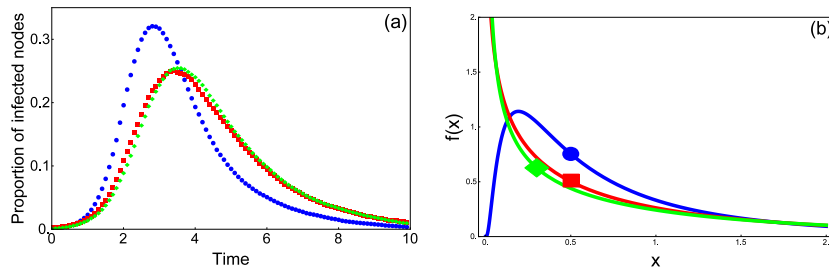
12

Figure 7. **(a)** Comparison of three epidemic curves after averaging explicit stochastic simulations with three different distributions of recovery times. The diamond/circle/square corresponds to Gamma$(0.5, 2)$, $\ln\mathcal{N}\left(\frac{\ln(3)}{2}, \sqrt{\ln(3))}\right)$ and Weibull$(0.72, 0.81)$ distributions, respectively. All three distributions have mean $m = 1$ and variance $v = 2$. **(b)** Probability density functions corresponding to the three distributions.

leads to higher reproduction number, and consequently the epidemic curve is characterised by faster initial growth rate and higher prevalence peak, see Figs. 1, 3 and 5. We note that our simulations were done only for homogeneous graphs, where the pairwise approximation works well [5]. The possible interplay of degree heterogeneity, or the clustering of the network, and the distribution of recovery times is an interesting future question.

For two-parameter distribution families, it is possible to regard $\mathcal{R}_0^p$ as a function depending on two variables, e.g. the mean m and variance v, see Figs. 2, 4, 6. Since our general final size relation (10) is monotone in $\mathcal{R}_0^p$, we conclude that smaller variance generates more infections.

It is important to observe that this statement, i.e. that smaller variance implies higher $\mathcal{R}_0^p$ is true only if we compare distributions from the same family. In Fig. 7, we compared three distributions from different families, each having $m = 1$ and $v = 2$. Besides the gamma and the lognormal distributions, for the sake of comparison we selected a third type of continuous distribution, namely Weibull distribution, which has been fitted to the infectious period for the recent ebola outbreak [4]. Fig. 7 illustrates that the mean and the variance of the recovery times alone are not able to determine the key characteristics of the epidemic curves, and a large variety of outbreaks can be generated from having the same mean and variance. This is especially the case in Fig. 7, where the gamma distributed infectious period leads to a very different epidemic, compared to that corresponding to the lognormally distributed infectious period, despite the mean and the variance are being identical. Therefore, in a real life situation, it is crucial

to estimate the empirical distribution of the infectious period as accurately as possible, since the mean and the variance alone do not provide enough information for accurate predictions.

References

1. N.T.J. Bailey. On estimating of latent and infectious periods of measles. i. families with two susceptibles only. *Biometrika*, 43:15–22, 1956.
2. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4):175–308, 2006.
3. M. Boguná, L. F. Lafuerza, R. Toral, and M. A. Serrano. Simulating non-markovian stochastic processes. *Phys. Rev. E*, 90(4):042108, 2014.
4. G. Chowell and H. Nishiura. Transmission dynamics and control of Ebola Virus Disease (EVD): a review. *BMC Medicine*, 196, 2014.
5. F. Cooper. Non-Markovian network epidemics, MA Thesis, 2013. *www.dtc.ox.ac.uk/people/13/cooperf/files/MA469ThesisFergusCooper.pdf*
6. E. Cator, R. Van de Bovenkamp, and P. Van Mieghem. Susceptible-infected-susceptible epidemics on networks with general infection and cure times. *Phys. Rev. E*, 87(6):062816, 2013.
7. L. Danon, A. P. Ford, T. House, C. P. Jewell, M. J. Keeling, G. O. Roberts, J. V. Ross, and M. C. Vernon. Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases*, 2011.
8. O. Diekmann, J. A. P. Heesterbeek, and J. A. J. Metz. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. *J. Math. Biol.* 28(4):365-382, 1990.
9. M. Eichner and K. Dietz. Transmission potential of smallpox: estimates based on detailed data from an outbreak. *Am. J. Epidemiol.*, 158:110–117, 2003.
10. S. Gubbins, S. Carpenter, M. Baylis, J. L. Wood, and P. S. Mellor. Assessing the risk of bluetongue to UK livestock: uncertainty and sensitivity analyses of a temperature-dependent model for the basic reproduction number. *Journal of the Royal Society Interface,* 5(20):363-371, 2008.
11. T. Hoffmann, M. A. Porter, and R. Lambiotte. Generalized master equations for non-Poisson dynamics on networks. *Phys. Rev. E*, 86(4):046102, 2012.
12. H.-H. Jo, J. I. Perotti, K. Kaski, and J. Kertész. Analytically solvable model of spreading dynamics with non-Poissonian processes. *Phys. Rev. X*, 4(1):011041, 2014.
13. M. J. Keeling and K. T. D. Eames. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307, 2005.
14. M. J. Keeling. The effects of local spatial structure on epidemiological invasions. *Proc. R. Soc. B*, 266(1421):859–867, 1999.
15. M. J. Keeling and B. T. Grenfell. Understanding the persistence of

14

measles: reconciling theory, simulation and observation. *Proc. R. Soc. B*, 269(1489):335–343, 2002.

16. E. Kenah and J. M. Robins. Second look at the spread of epidemics on networks. *Physical Review E*, 76(3):036113, 2007.

17. I. Z. Kiss, G. Röst, and Zs. Vizi. Pairwise-like models for non-Markovian epidemics on networks. *Phys. Rev. Letters, in press, arXiv preprint arXiv:1504.04639*, 2015.

18. A. L. Lloyd. Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theor. Pop. Biol.*, 60(1):59–71, 2001.

19. B. Min, K. I. Goh, and I. M. Kim. Suppression of epidemic outbreaks with heavy-tailed contact dynamics. *EPL (Europhysics Letters)*, 103(5):50002, 2013.

20. H. Nishiura and M. Eichner. Infectiousness of smallpox relative to disease age: estimates based on transmission network and incubation period. *Epidemiol. Infect.*, 135:1145–1150, 2007.

21. M. E. J. Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.

22. M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

23. R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *arXiv preprint arXiv:1408.2701*, 2014.

24. P. Van Mieghem and R. Van de Bovenkamp. Non-Markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks. *Phys. Rev. Lett.*, 110(10):108701, 2013.

25. R. R. Wilkinson and K. J. Sharkey. Message passing and moment closure for susceptible-infected-recovered epidemics on finite networks. *Physical Review E*, 89(2):022808, 2014.

26. Y. Yang, M. E. Halloran, J. D. Sugimoto and I. M. Longini Jr. Detecting human-to-human transmission of avian influenza A (H5N1). *Emerging Infectious Diseases*, 13(9), 1348, 2007.