# Variance of Infectious Periods and Reproduction Numbers for Network Epidemics with Non-Markovian Recovery

Gergely Röst, István Z. Kiss and Zsolt Vizi

**Abstract** For a recently derived pairwise model of network epidemics with non-Markovian recovery, we prove that under some mild technical conditions on the distribution of the infectious periods, smaller variance in the recovery time leads to higher reproduction number when the mean infectious period is fixed.

## 1 Introduction

Networks provide a useful paradigm to incorporate contact patterns and various heterogeneities within a population [12, 8]. The basic ingredients of such models are nodes and links, usually representing individuals and the contacts between them, but they may represent also groups of individuals (such as the population at some geographic location), and the connectedness of these groups (such as transportation routes [4, 7]). In simple disease outbreak models, the status of an individual can be susceptible ($S$), infected ($I$) or recovered ($R$). A key parameter associated with most epidemic models is the basic reproduction number (denoted by $\mathscr{R}_0$), which denotes the expected number of secondary infections generated by a typical infected individual introduced into a fully susceptible population [2]. The reproduction number is also a threshold quantity: if $\mathscr{R}_0 < 1$ the epidemic will die out, while if $\mathscr{R}_0 > 1$ the disease will spread. Another important measure of epidemic severity is the final epidemic size, which is the total number of individuals who become infected during

Gergely Röst
Bolyai Institute, University of Szeged, e-mail: rost@math.u-szeged.hu

István Z. Kiss
Department of Mathematics, University of Sussex, e-mail: i.z.kiss@sussex.ac.uk

Zsolt Vizi
Bolyai Institute, University of Szeged, e-mail: vizi@math.u-szeged.hu
and Robert Bosch Ltd., Engineering Centre of Budapest

the time course of the epidemic. These two quantities are often connected via the so-called final size relation.

Pairwise models have been successfully used to approximate stochastic epidemics on networks and represent an improvement on compartmental models. The former are formulated in terms of the expected values for the number of susceptible ($[S]$), infected ($[I]$) and recovered ($[R]$) nodes, which depend on the expected values of ($SS$) pairs ($[SS]$) and ($SI$) pairs ($[SI]$). Introducing the usual notations

- $[X](t)$ for the expected number of nodes in state $X$ at time $t$,
- $[XY](t)$ for the expected number of links connecting a node in state $X$ to another in state $Y$, and
- $[XYZ](t)$ for the expected number of triplets in state $X - Y - Z$,

where, $X, Y, Z \in \{S, I, R\}$, and by summing up all possible transitions, the pairwise model reads as

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

where $\tau$ is the per contact infection rate and $\gamma$ is the recovery rate. Here $[S] + [I] + [R] = N$ is the total number of nodes in the network, and only those equations are listed which are necessary to derive a complete self-consistent system. The equations for links contain triplets, thus we have to break the dependence on higher order terms to obtain a closed system. The closure approximation formula $[XSY] = \frac{n-1}{n} \frac{[XS][SY]}{[S]}$, where $n$ is the average number of links per node, leads to the self-consistent system [3]

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

Closing at the level of pairs with the approximation $[XY] = n[X]\frac{[Y]}{N}$, one obtains the so called mean-field model (or compartmental model)

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

with basic reproduction number

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

where, $\mathbb{E}(\mathscr{I}) = 1/\gamma$ is the expected infectious period. There are many results for the Markovian case [1, 3, 5], for example, the final epidemic size is given by

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

where $[S]_0$ is the number of susceptible individuals at time $t = 0$ and $s_\infty = [S]_\infty/[S]_0$, where $[S](\infty) = [S]_\infty$.

## 2 Non-Markovian Recovery

The Markovianity of the recovery process is a strong simplifying assumption. For many epidemics, the infectious period has great importance and it is measured empirically. Recently, pairwise approximations of the SIR dynamics with non-Markovian recovery have been derived, see [6, 11, 9, 10]. In the special of fixed recovery time $\sigma$, the mean-field model is given by

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

while the pairwise model turned out to be [6]

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

Both systems are now delay differential equations rather than ordinary differential equations, as is the case for Markovian epidemics. Considering a general distribution for the recovery period, the pairwise model can be formulated as a system of integro-differential equations [10, 11], which we omit here. In [6], the following final epidemic size relation has been derived:

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

## 3 The Pairwise Reproduction Number and Recovery Times

In [6], a newly introduced basic reproduction-like number is defined as $\mathscr{R}_0^p = \frac{n-1}{N}\left(1 - e^{-\tau\sigma}\right)[S]_0$, which appears also in (8). It has also been shown, that for arbitrary infectious periods, the basic reproduction number of the pairwise model is

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

where $\mathscr{L}[\cdot]$ is the Laplace transform and $f_{\mathscr{I}}$ is the probability density function of the recovery process given by the random variable $\mathscr{I}$. Numerical tests and analytical results have both confirmed that, in general, the following implicit relation for the final epidemic size holds

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

Notice that while $\mathscr{R}_0$ depends on the expected value only, see (4), the pairwise reproduction number (9) uses the complete density function, thus the average length of the infectious period does not determine exactly the reproduction number. As a consequence, for an epidemic we have to know as precisely as possible the shape of the distribution. We shall analyze how the basic reproduction number (9), which is not only an epidemic threshold but also determines the final size via (10), depends on the variance of the recovery time distribution. In [9], using gamma, lognormal and uniform distributions we showed that once the mean infectious period is fixed, smaller variance in the infectious period gives a higher reproduction number and consequently a more severe epidemic. Next we generalize this result without restricting ourselves to special distributions.

## 4 Relationship Between the Variance and the Reproduction Number

In this section we give some simple conditions which may guarantee that smaller variance induces higher pairwise reproduction number. We consider a random variable $\mathscr{I}$ corresponding to recovery times with probability density functions $f_{\mathscr{I}}(t)$, cumulative distribution function $F_{\mathscr{I}}(t) = \int_0^t f_{\mathscr{I}}(s)ds$ and we shall use the integral function of the CDF $\mathscr{F}_{\mathscr{I}}(t) := \int_0^t F_{\mathscr{I}}(s)ds$. Clearly, $\frac{d^2}{dt^2}\mathscr{F}_{\mathscr{I}}(t) = \frac{d}{dt}F_{\mathscr{I}}(t) = f_{\mathscr{I}}(t)$. Moreover, $F_{\mathscr{I}}(0) = \mathscr{F}_{\mathscr{I}}(0) = 0$.

**Theorem 1.** *Consider two random variables $\mathscr{I}_1$ and $\mathscr{I}_2$ such that*

$$\mathbb{E}(\mathscr{I}_1) = \mathbb{E}(\mathscr{I}_2) < \infty, \tag{11}$$

*and*

$$\mathrm{Var}(\mathscr{I}_1) < \mathrm{Var}(\mathscr{I}_2) < \infty. \tag{12}$$

*Assume that*

$$\lim_{t\to\infty} t^3 f_{\mathscr{I}_j}(t) = 0, \quad j \in \{1,2\}, \tag{13}$$

*and for all $t > 0$,*

$$\mathscr{F}_{\mathscr{I}_1}(t) \neq \mathscr{F}_{\mathscr{I}_2}(t) \tag{14}$$

*holds. If $\mathscr{I}_1$ and $\mathscr{I}_2$ represent the recovery time distribution, then for the corresponding reproduction numbers the relation $\mathscr{R}_{0,\mathscr{I}_1}^p > \mathscr{R}_{0,\mathscr{I}_2}^p$ holds.*

*Proof.* Using assumption (11), we deduce

$$
\begin{aligned}
\int_0^\infty t\left(f_{\mathscr{I}_1}(t) - f_{\mathscr{I}_2}(t)\right) &= \left[t(F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t))\right]_0^\infty - \int_0^\infty (F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t))dt \\
&= \lim_{t\to\infty} t(F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t)) - \left[\mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t)\right]_0^\infty \\
&\overset{[*]}{=} -\lim_{t\to\infty}(\mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t)) = 0
\end{aligned}
$$

thus

$$\lim_{t\to\infty}(\mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t)) = 0. \tag{15}$$

To see $[*]$, i.e. $\lim_{t\to\infty} t(F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t)) = 0$, we need some algebraic manipulations:

$$
\begin{aligned}
\lim_{t\to\infty} t(F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t)) &= \lim_{t\to\infty} \frac{F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t)}{\frac{1}{t}} \overset{\text{L'H}}{=} \lim_{t\to\infty} \frac{f_{\mathscr{I}_1}(t) - f_{\mathscr{I}_2}(t)}{-\frac{1}{t^2}} \\
&= -\lim_{t\to\infty} t^2(f_{\mathscr{I}_1}(t) - f_{\mathscr{I}_2}(t)) \overset{(13)}{=} 0,
\end{aligned}
$$

where L'H refers to the L'Hospital rule. From assumption (12), we have

$$
\begin{aligned}
\mathrm{Var}(\mathscr{I}_1) = \mathbb{E}(\mathscr{I}_1^2) - (\mathbb{E}(\mathscr{I}_1))^2 &< \mathbb{E}(\mathscr{I}_2^2) - (\mathbb{E}(\mathscr{I}_2))^2 = \mathrm{Var}(\mathscr{I}_2) \\
&\overset{(11)}{\Rightarrow} \mathbb{E}(\mathscr{I}_1^2) < \mathbb{E}(\mathscr{I}_2^2).
\end{aligned}
$$

or equivalently $\int_0^\infty t^2(f_{\mathscr{I}_1} - f_{\mathscr{I}_2})dt < 0$. We can carry out some calculation on the left-hand side of this inequality:

$$
\begin{aligned}
\int_0^\infty t^2(f_{\mathscr{I}_1} - f_{\mathscr{I}_2})dt &= [t^2(F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t))]_0^\infty - 2\int_0^\infty t(F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t))dt \\
&= \lim_{t\to\infty} t^2(F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t)) - 2[t(\mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t))]_0^\infty \\
&\quad + 2\int_0^\infty \mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t)dt \\
&\overset{[**]}{=} -2\lim_{t\to\infty} t(\mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t)) + 2\int_0^\infty \mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t)dt \\
&\overset{[**]}{=} 2\int_0^\infty \mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t)dt,
\end{aligned}
$$

consequently

$$\int_0^\infty \mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t)dt < 0 \tag{16}$$

To prove [$**$], i.e. $\lim_{t\to\infty} t^2(F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t)) = \lim_{t\to\infty} t(\mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t)) = 0$, we have

$$
\begin{aligned}
\lim_{t\to\infty} t(\mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t)) &= \lim_{t\to\infty} \frac{\mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t)}{\frac{1}{t}} \overset{\text{L'H}}{=} \lim_{t\to\infty} \frac{F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t)}{-\frac{1}{t^2}} \\
&= -\lim_{t\to\infty} t^2(F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t)) \\
&\overset{\text{L'H}}{=} \lim_{t\to\infty} \frac{f_{\mathscr{I}_1}(t) - f_{\mathscr{I}_2}(t)}{\frac{2}{t^3}} = \frac{1}{2}\lim_{t\to\infty} t^3(f_{\mathscr{I}_1}(t) - f_{\mathscr{I}_2}(t)) \\
&\overset{(13)}{=} 0.
\end{aligned}
$$

Since $F_{\mathscr{I}}(t) \geq 0, t \geq 0$ and monotone increasing, the integral function of CDF $\mathscr{F}_{\mathscr{I}}(t)$ is monotone increasing and convex. Using (14) and (16), we obtain

$$\mathscr{F}_{\mathscr{I}_1}(t) < \mathscr{F}_{\mathscr{I}_2}(t), \tag{17}$$

for all $t > 0$. Clearly, for $\mathscr{R}^p_{0,\mathscr{I}_1} > \mathscr{R}^p_{0,\mathscr{I}_2}$, it is enough to prove, that $\mathscr{L}[f_{\mathscr{I}_1}](\tau) < \mathscr{L}[f_{\mathscr{I}_2}](\tau)$, i.e. $\int_0^\infty e^{-\tau t}(f_{\mathscr{I}_1}(t) - f_{\mathscr{I}_2}(t))dt < 0$. First, we perform some algebraic manipulation on the left-hand side:

$$
\begin{aligned}
\int_0^\infty e^{-\tau t}(f_{\mathscr{I}_1}(t) - f_{\mathscr{I}_2}(t))dt &= [e^{-\tau t}(F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t))]_0^\infty \\
&\quad + \tau\int_0^\infty e^{-\tau t}(F_{\mathscr{I}_1}(t) - F_{\mathscr{I}_2}(t))dt \\
&= \tau[e^{-\tau t}(\mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t))]_0^\infty \\
&\quad + \tau^2\int_0^\infty e^{-\tau t}(\mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t))dt \\
&\overset{(15)}{=} \tau^2\int_0^\infty e^{-\tau t}(\mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t))dt.
\end{aligned}
$$

In conclusion, we have

$$\tau^2\int_0^\infty e^{-\tau t}(\mathscr{F}_{\mathscr{I}_1}(t) - \mathscr{F}_{\mathscr{I}_2}(t))dt \overset{(17)}{<} 0,$$

therefore $\mathscr{L}[f_{\mathscr{I}_1}](\tau) < \mathscr{L}[f_{\mathscr{I}_2}](\tau)$, which gives $\mathscr{R}^p_{0,\mathscr{I}_1} > \mathscr{R}^p_{0,\mathscr{I}_2}$.

## 5 Conclusion

Our previous works already indicated that for pairwise models not only the mean, but higher order properties of the distribution of the recovery times have an impact on the outcome of the epidemic. We derived useful threshold quantities for non-Markovian recovery in [6]. In [9], we showed that for particular distribution families (typically two parameter families such as gamma, lognormal, and uniform distribution), smaller variance leads to higher reproduction number within the same family when the mean is fixed. Our new result in this study allows as to make comparisons between distributions of different kinds. To show the usefulness of Theorem 1, as an example, we consider $\mathscr{I}_1 \sim \text{Exp}(\gamma)$ and $\mathscr{I}_2 \sim \text{Fixed}\left(\frac{1}{\gamma}\right)$, i.e. $f_{\mathscr{I}_1}(t) = \gamma e^{-\gamma t}, t \geq 0$ and $f_{\mathscr{I}_2}(t) = \delta\left(t - \frac{1}{\gamma}\right)$, where $\delta(t)$ denotes the Dirac delta function. Clearly, we obtain $\mathscr{F}_{\mathscr{I}_1}(t) = t + \frac{1}{\gamma}e^{-\gamma t} - \frac{1}{\gamma}$ and $\mathscr{F}_{\mathscr{I}_2}(t) = t - \frac{1}{\gamma}$, thus there is no $t_0 > 0$, such that $\mathscr{F}_{\mathscr{I}_1}(t_0) = \mathscr{F}_{\mathscr{I}_2}(t_0)$. Since $\mathbb{E}(\mathscr{I}_1) = \mathbb{E}(\mathscr{I}_2) = \frac{1}{\gamma}$, $\frac{1}{\gamma^2} = \text{Var}(\mathscr{I}_1) > \text{Var}(\mathscr{I}_2) = 0$ and the other conditions of Theorem 1 are satisfied, we find $\mathscr{R}^p_{0,\mathscr{I}_1} < \mathscr{R}^p_{0,\mathscr{I}_2}$.

## References

1. Carlos, L., Juher, D., and Saldaña, J.: On the early epidemic dynamics for pairwise models. J. Theor. Biol. **352**, 71–81 (2014)
2. Diekmann, O., Heesterbeek, J.A.P., and Metz, J.A.J.: 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)
3. Keeling, M.J.: The effects of local spatial structure on epidemiological invasions. Proc. R. Soc. Lond. B **266**, 859–867 (1999)
4. Knipl, D., Röst, G.: Large number of endemic equilibria for disease transmission models in patchy environment. Math. Biosci. **258**, 201–222 (2014)
5. Kiss, I.Z., Miller, J.C., and Simon, L.P.: Mathematics of Epidemics on Networks – From Exact to Approximate Models. Springer (2017)
6. Kiss, I.Z., Röst, G., and Vizi, Zs.: Generalization of pairwise models to non- Markovian epidemics on networks. Phys. Rev. Lett., **115** (7) 078701 (2015)
7. Nakata Y., Röst G.: Global analysis for spread of infectious diseases via transportation networks. J. Math. Biol., **70** (6), 1411–1456 (2015)
8. Newman, M.E.J.: Spread of epidemic disease on networks. Phys. Rev. E, **66** (1) 016128 (2002)
9. Röst, G., Vizi, Zs., and Kiss, I. Z.: Impact of non-Markovian recovery on network epidemics, in: Mondaini RP (ed.) BIOMAT 2015, World Scientific, 40–53 (2016)
10. Röst, G., Vizi, Zs., and Kiss, I. Z.: Pairwise approximation for SIR type network epidemics with non-Markovian recovery. arXiv:1605.02933 (2016)
11. Wilkinson, R.R., Ball, F.G., and Sharkey, K.J.: The relationships between message passing, pairwise, Kermack-McKendrick and stochastic SIR epidemic models. arXiv:1605.03555 (2016)
12. Pastor-Satorras, R., Castellano, C., Van Mieghem, P., and Vespignani, A.: Epidemic processes in complex networks. Rev. Mod. Phys. **87** (3) 925 (2015)