Simulation

This is a translated version based on [Todd's notes](#).
We are able to describe the result with a short ODE and being able to solve it numerically together with ODE system of MSV framework.

We would like to compute:

$$p(t) = \mathbb{P}\{\text{a node infected at time } t \text{ infects a randomly given neighbour}\}$$

- As in MSV framework, we assume infection of neighbors are independent, thus the number of infected neighbors follows binomial distribution. Follow the idea of Zhao2 result/derivation:
  - For newly infected vertices, we condition on that the neighbor must not be the infector of the focal vertex, i.e. the edge has not transmitted the infection yet s.t.

$$\mu(t) = \frac{p(t)}{\phi(t)}$$

  - Then, correspond to $\mathcal{R}_c^*$, we have

$$\mathcal{R}_c(t) = \mu(t) \times (\mathbb{E}[K_I^*] - 1) = p(t) \times \frac{G_p''(\phi(t))}{G_p'(\phi(t))}$$

- (**Wrong here**)Initially at $t = 0$, there should be no competing infection among neighbors, all neighbors of focal infected node can only be infected by the focal node.
  - This leads to

$$p(0) = \tau = \frac{\beta}{\beta + \gamma}$$

    as initial value of $p(t)$.
  - Furthermore, this agree with the initial value for $\mathcal{R}_c$, s.t.

$$\mathcal{R}_c(0) = \mathcal{R}_c^*(0) = \mathcal{R}_{c,0} = \frac{\beta}{\beta + \gamma} \times \frac{G_p''(1)}{\delta}$$

To find $p(t)$ in the stochastic process, we observe that the focal infected node will infect its neighbor:

- if the focal node makes an infectious contact with the neighbor under rate β before the focal node's recovery under rate γ
- $\textbf{AND}$ the neighbor is not infected (by the neighbor's neighbor other than the focal) at the time of contact by focal.
  With the MSV framework, the probability that a randomly chosen neighbor of the focal

node is not infected at time t is

$$\phi_S(t) = G_q(\phi(t)) = \frac{G'_p(\phi(t))}{\delta}$$

In the random events, lets set the following random variables for time:

- T_r: the time after infection t that the focal infected node recovers. Based on the recovery rate γ and exponential distribution, we have

$$\mathbb{P}(T_r > s) = e^{-\gamma s}$$

- T_c: the time after infection t that the focal node makes it infectious contact with the neighbor through the edge connecting them. Based on the infection rate β and exponential distribution, we have

$$\mathbb{P}(T_c > s) = e^{-\beta s}$$

- T_n: the first time that the neighbor has an infectious contact from one of its other neighbors than the focal node. We further have

$$\mathbb{P}(T_n > t + s) = \phi_S(t + s)$$

With these RVs, we can interpret $p(t)$ in the following probability:

$$p(t) = \mathbb{P}\{t + T_c < (t + T_r) \wedge T_n\}$$

where $t + T_c < (t + T_r) \wedge T_n \;\Leftrightarrow \min((t + T_r), T_n)$.

Since T_r and T_n are independent, we can further derive:

$$\begin{aligned} \mathbb{P}\{(t + T_r) \wedge T_n > t + u\} &= \mathbb{P}\{T_r > u\}\mathbb{P}\{T_n > t + u\} \\ &= e^{-\gamma u}\phi_S(t + u) \end{aligned}$$

Then we can rewrite the expression for $p(t)$ based on law of total probability:

$$\begin{aligned} p(t) &= \mathbb{P}\{t + T_c < (t + T_r) \wedge T_n\} \\ &= \int_0^\infty \mathbb{P}\{t + u < (t + T_r) \wedge T_n | t + T_c = t + u\} \times \mathbb{P}\{t + T_c = t + u\} du \\ &= \int_0^\infty \mathbb{P}\{t + u < (t + T_r) \wedge T_n | T_c = u\} \times \mathbb{P}\{T_c = u\} du \\ &= \int_0^\infty \mathbb{P}\{t + u < (t + T_r) \wedge T_n\} \times \mathbb{P}\{T_c = u\} du \quad \text{as } T_n, T_c, T_r \text{ are independent} \\ &= \int_0^\infty [e^{-\gamma u}\phi_S(t + u)] \times [\beta e^{-\beta u}] du \\ &= \int_0^\infty \beta e^{-(\beta+\gamma)u}\phi_S(t + u) du \end{aligned}$$

and we have

$$
\begin{aligned}
\frac{d}{dt}p(t) &= \frac{d}{dt}\int_0^\infty \beta e^{-(\beta+\gamma)u}\phi_S(t+u)du \\
&= \int_0^\infty \beta e^{-(\beta+\gamma)u} \times [\frac{d}{dt}\phi_S(t+u)]du \\
&= \int_0^\infty \beta e^{-(\beta+\gamma)u} \times [\dot{\phi}_S(t+u)]du \\
&= [\beta e^{-(\beta+\gamma)u}\phi_S(t+u)]\Big|_{u=0}^{\infty} - \int_0^\infty -(\beta+\gamma)\beta e^{-(\beta+\gamma)u}\phi_S(t+u)du \quad \text{(I.B.P.)} \\
&= -\beta\phi_S(t) + (\beta+\gamma)\int_0^\infty \beta e^{-(\beta+\gamma)u}\phi_S(t+u)du \\
&= -\beta\phi_S(t) + (\beta+\gamma)p(t)
\end{aligned}
$$

as a ODE of $p(t)$.

At $t = 0$, we expect to have $\phi_S(0) = 1$ and $p(0) = \frac{\beta}{\beta+\gamma}$, so take these initial value into the ODE gives us

$$
\frac{d}{dt}p(t) = -\beta + (\beta+\gamma) \times \frac{\beta}{\beta+\gamma} = 0
$$

which agree with our expectation.

Sign Problem

This notes fixed some typo in Todd's original notes and the final ODE agree with the original one. I review the derivation myself twice and it seems correct.

But we expect $p(t)$ be a probability and monotonically decreasing from its initial value since the following two factor:

- ϕ_S is decreasing as the infection spread out.
- We have more competing infection happens, i.e. it is more likely to have $T_n < t + T_c$
  However, in numeric solving, this ODE for $p(t)$ leads to increasing $p(t)$ and getting larger than 1.

See R_c-SignIssue.R for numeric solutions `CM_P`, the ODE is implemented at line `#107`.

Just an observation: if alter the all signs in the ODE, the results of $p(t)$ seems to be much more reasonable.
Result: the sign is correct, but the problem is no an initial value problem, i.e. $p(0) \neq \frac{\beta}{\beta+\gamma}$. For details, see next chapter.

Reverse ODE

It turns out that the $p(t)$ value should not be a initial value problem, but a final value problem.

$$
\frac{d}{dt}p(t) = -\beta\phi_S(t) + (\beta+\gamma)p(t)
$$

when $t \to +\infty$, $\frac{d}{dt}p(\infty) \to 0$ as the system reaching its equilibrium state.

For the current system, it means the end of outbreak where no active infection exist in the population, thus no competing infection for any newly introduced infection vertex.
Therefore, at the end of equilibrium state we should have:

$$\mathcal{R}_c(\infty) = \mathcal{R}_c^*(\infty)$$

which gives us:

$$p(\infty) = \frac{\beta}{\beta+\gamma}\phi_S(\infty) = \frac{\beta}{\beta+\gamma}\frac{G_p'(\phi(\infty))}{\delta}$$

and we have know from MSV frame work that

$$\phi(\infty) = \frac{\gamma}{\beta+\gamma} + \frac{\beta}{\beta+\gamma}\frac{G_p'(\phi(\infty))}{\delta}$$

So $p(\infty) = \phi(\infty) - \frac{\gamma}{\beta+\gamma}$.

For the initial status, even with very small initial infection proportion, we could not easily assume

$$\mathcal{R}_c^*(0) = \mathcal{R}_c(0) \Leftrightarrow p(0) = \frac{\beta}{\beta+\gamma}$$

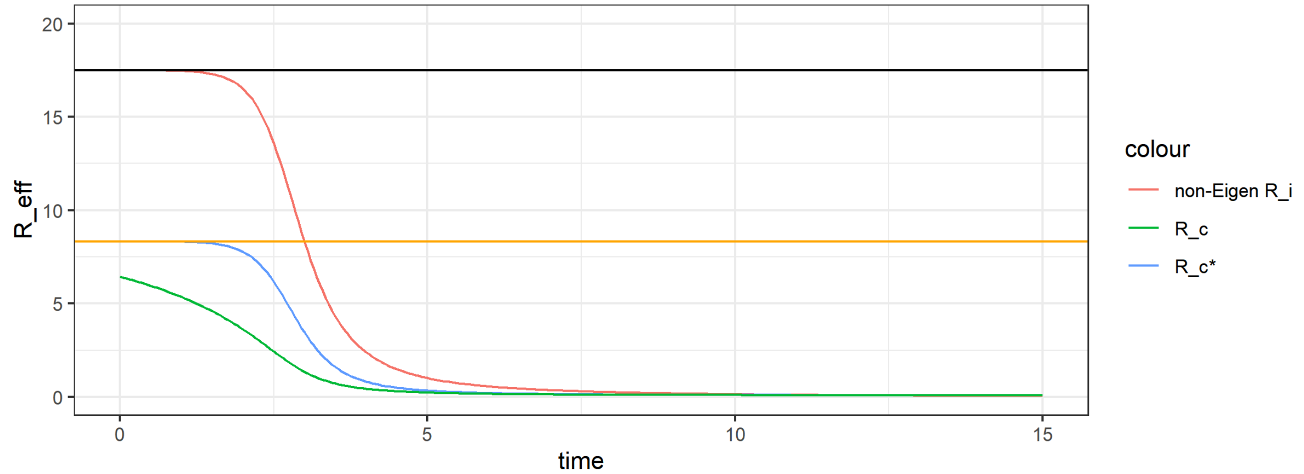
unless the network size N is large enough.
This is because future status will affect $p(t)$ by definition.
So at the very beginning, $p(0)$ might still be less than $\frac{\beta}{\beta+\gamma}$, if:

- γ is small(i.e. recovery time T_r is long)
- And/or N is not large enough (i.e. loops are rare enough for the whole outbreak) such that competition of infection is still affect $p(0)$ and $\mathcal{R}_c(0)$ value.
  More specifically, even if competing of infection has low probability at the $t = 0$, it still affect $p(0)$ as a lot of infection events happens even before the first infected individual recovers.

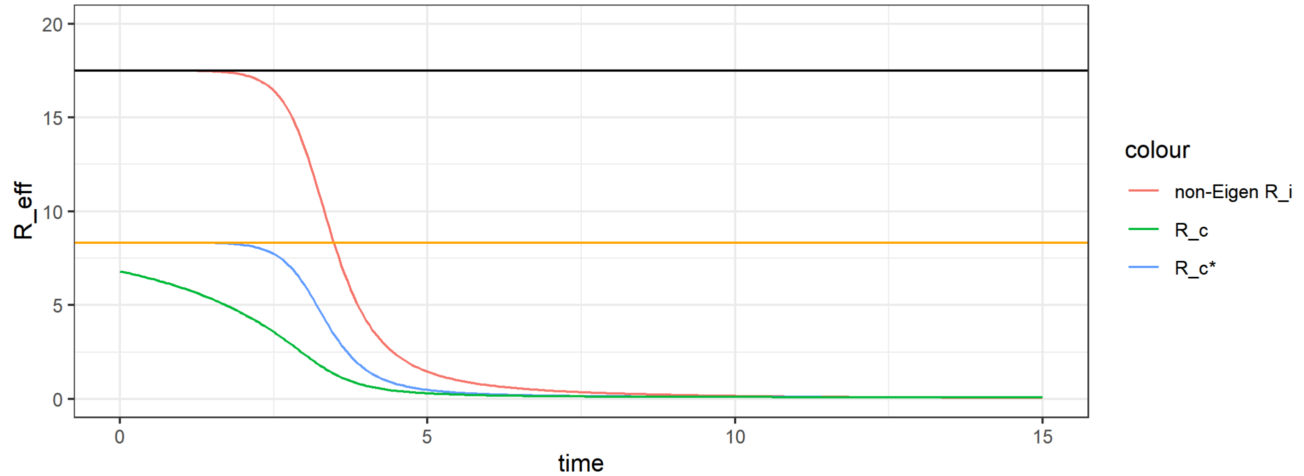
As a result, with the numerical simulation of the MSV dynamic, one can reversely simulate the $p(t)$ and the ODE, starting from some $p(t_f)$ as final value at a known time t_f close to $p(\infty)$, to get the $p(t)$ simulation and initial value $p(0)$.
Further we can calculate corresponding $\mathcal{R}_c$ and comparing with $\mathcal{R}_c^*$.

This difference between these two case-wise effective reproduction number is decreasing as γ and N increase:

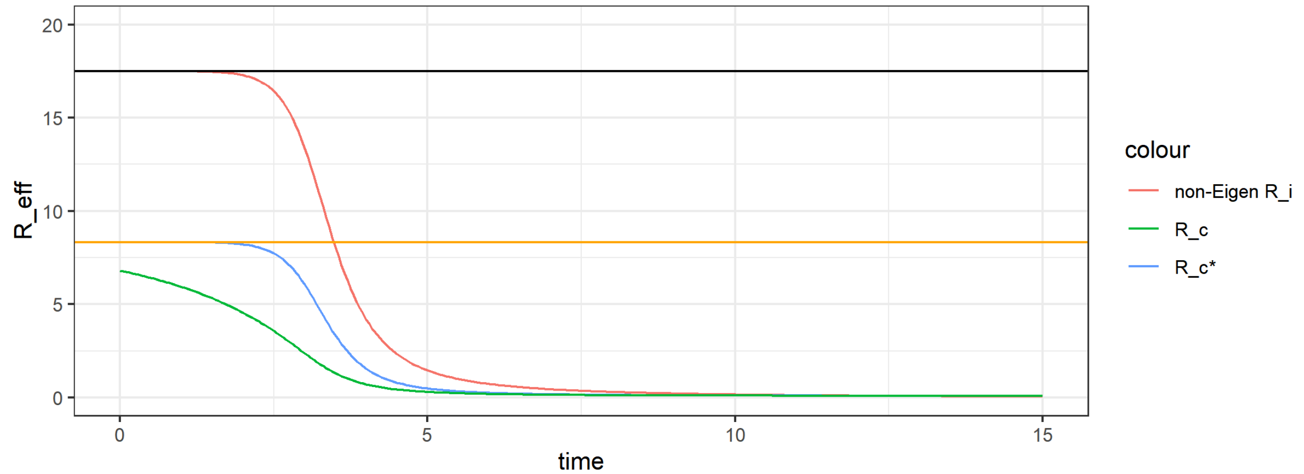
$\mathcal{R}_c(0) = 6.43, \mathcal{R}_c^*(0) = 8.33$ for $N = 50,000, \gamma = 0.20, \beta = 0.25, I_0 = 1$



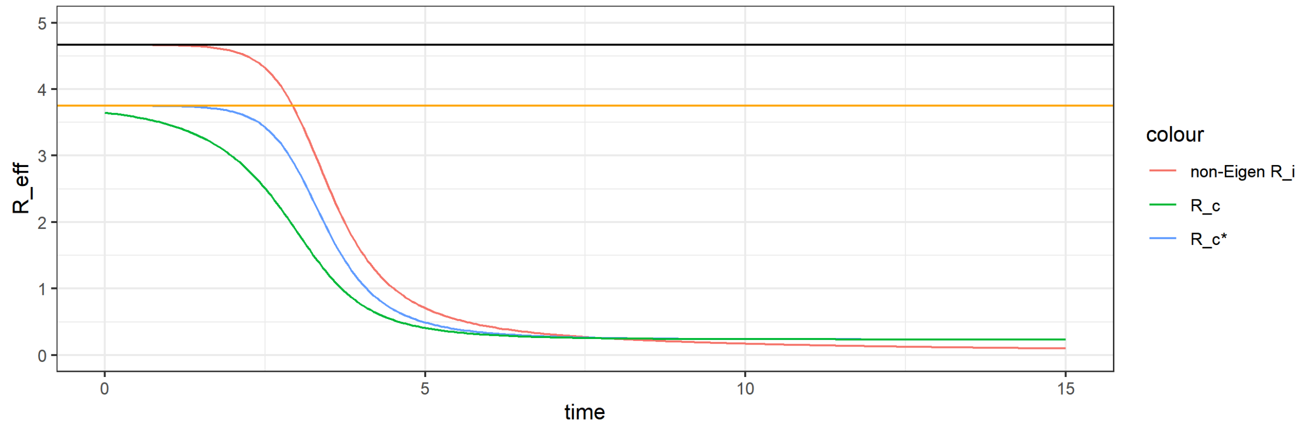
$\mathcal{R}_c(0) = 6.80, \mathcal{R}_c^*(0) = 8.33$ for $N = 250,000, \gamma = 0.20, \beta = 0.25, I_0 = 1$



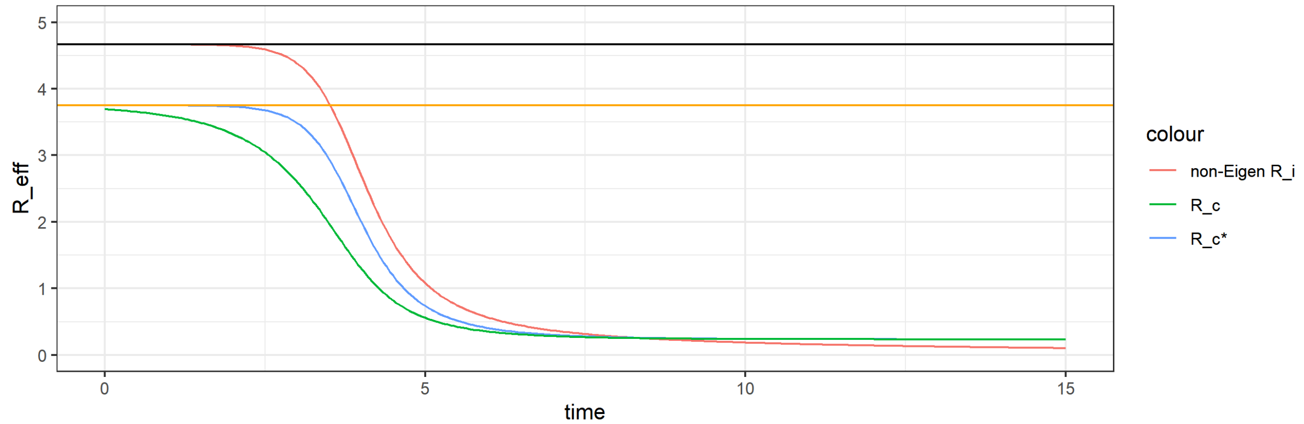
$\mathcal{R}_c(0) = 6.93, \mathcal{R}_c^*(0) = 8.33$ for $N = 500,000, \gamma = 0.20, \beta = 0.25, I_0 = 1$



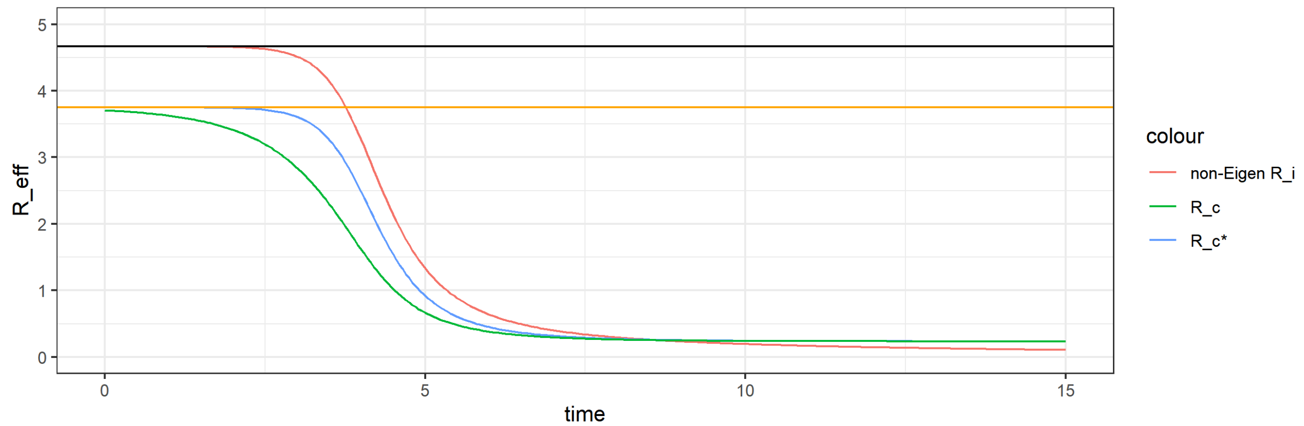
$\mathcal{R}_c(0) = 3.65, \mathcal{R}_c^*(0) = 3.75$ for $N = 50,000, \gamma = 0.75, \beta = 0.25, I_0 = 1$



$\mathcal{R}_c(0) = 3.69, \mathcal{R}_c^*(0) = 3.75$ for $N = 250,000, \gamma = 0.75, \beta = 0.25, I_0 = 1$



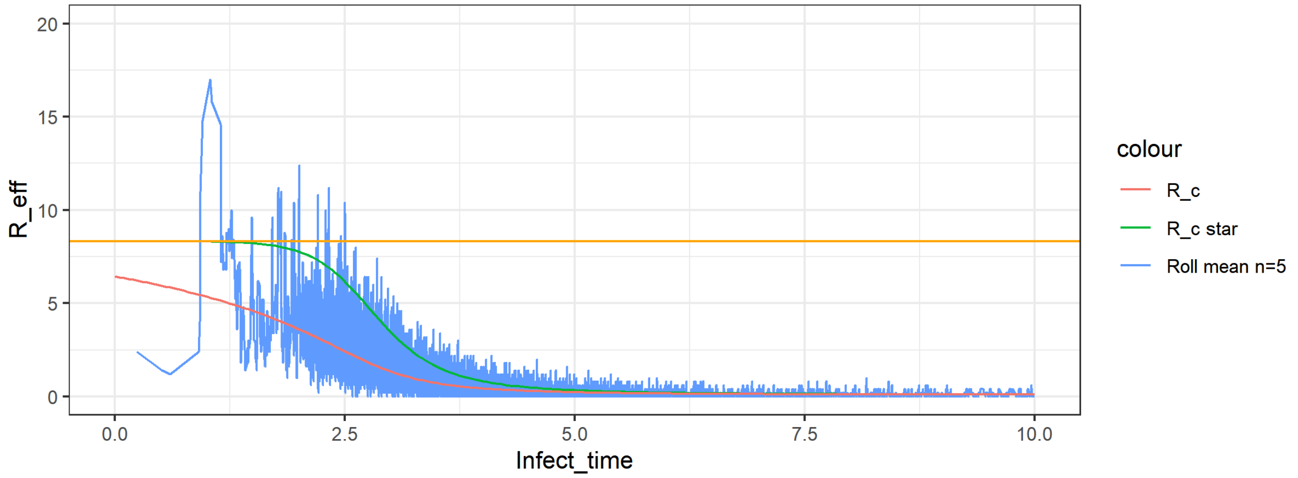
$\mathcal{R}_c(0) = 3.70, \mathcal{R}_c^*(0) = 3.75$ for $N = 500,000, \gamma = 0.75, \beta = 0.25, I_0 = 1$



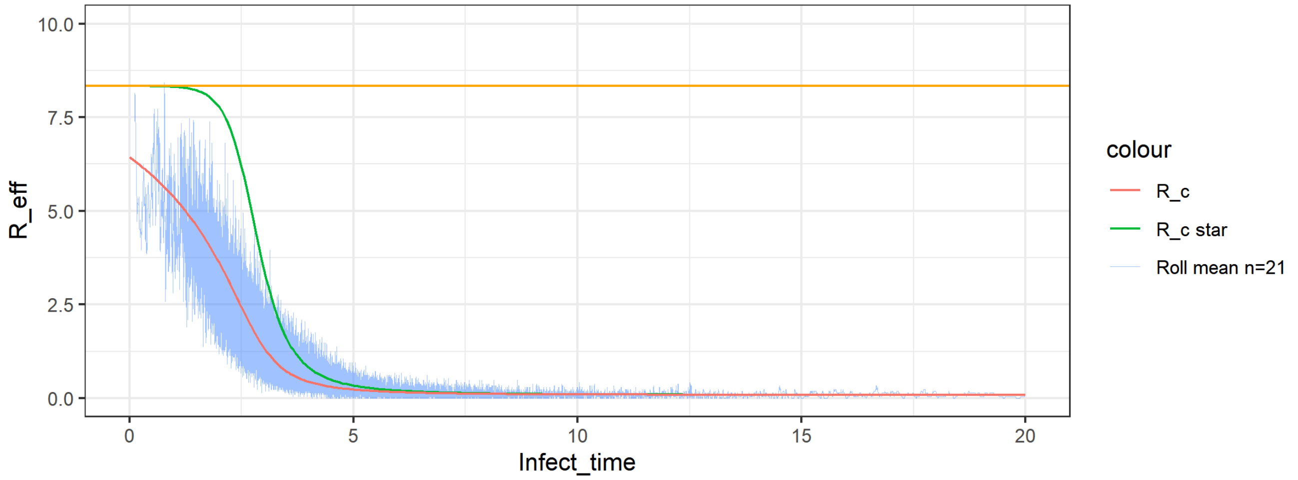
To further verify the $\mathcal{R}_c(t)$ and $p(t)$ result, I comparing the curves with the Gillespie simulation on network.

Simulation for $\gamma = 0.2$ case

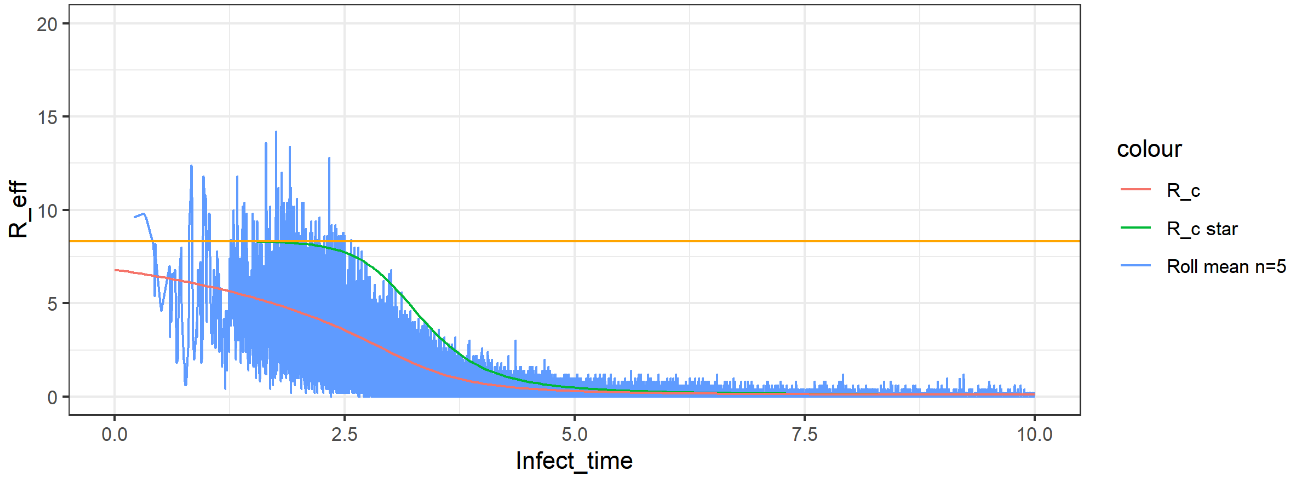
For $N = 50,000, \gamma = 0.20, \beta = 0.25, I_0 = 1$: 1 simulation



Timely-overlapped results for 16 simulation: (4 case with large phase shifting removed for 20 random runs)

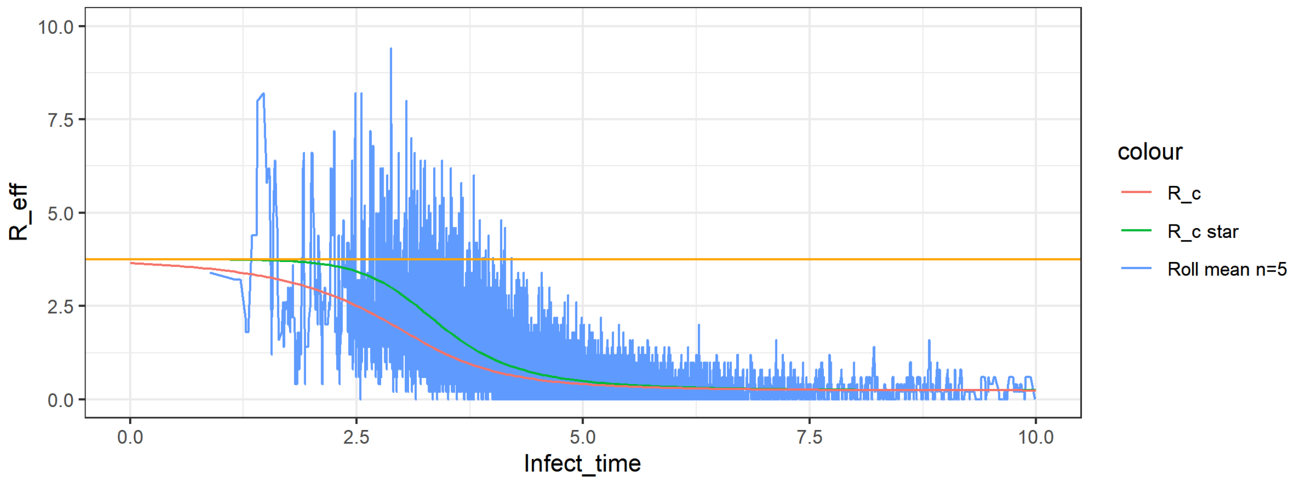


For larger network size , 1 simulation due to time consumption:

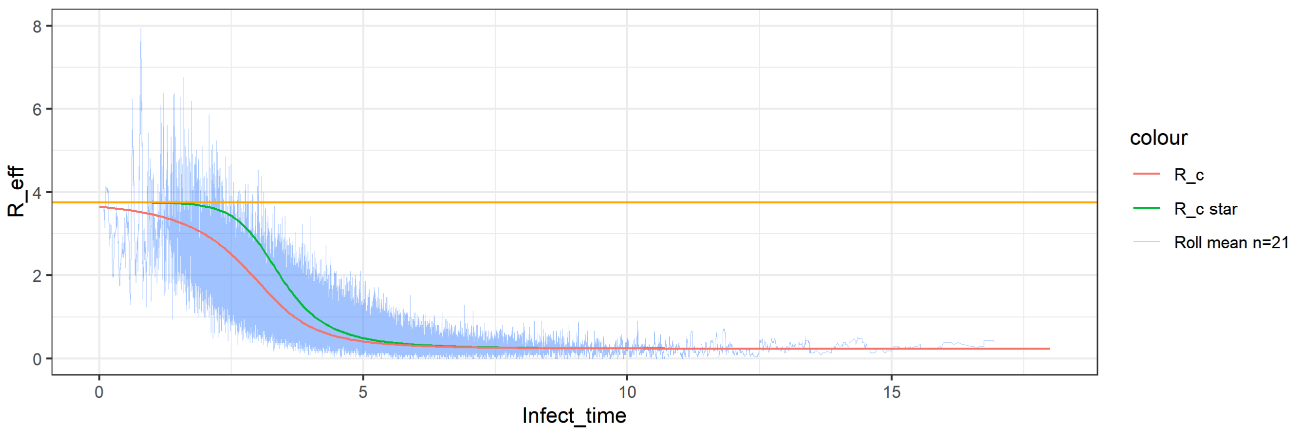


Simulation for $\gamma = 0.75$ case

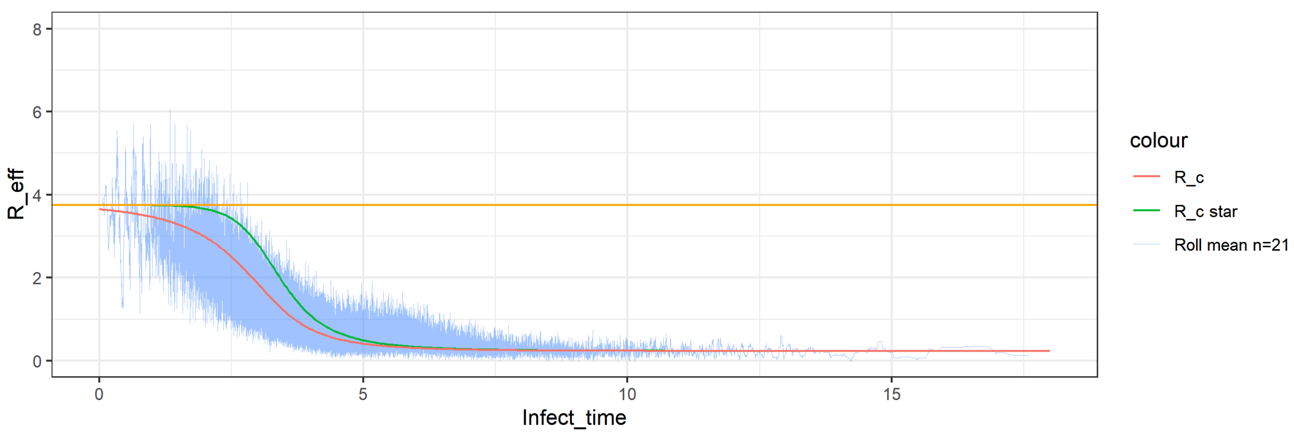
For $N = 50,000, \gamma = 0.75, \beta = 0.25, I_0 = 1$: 1 simulation



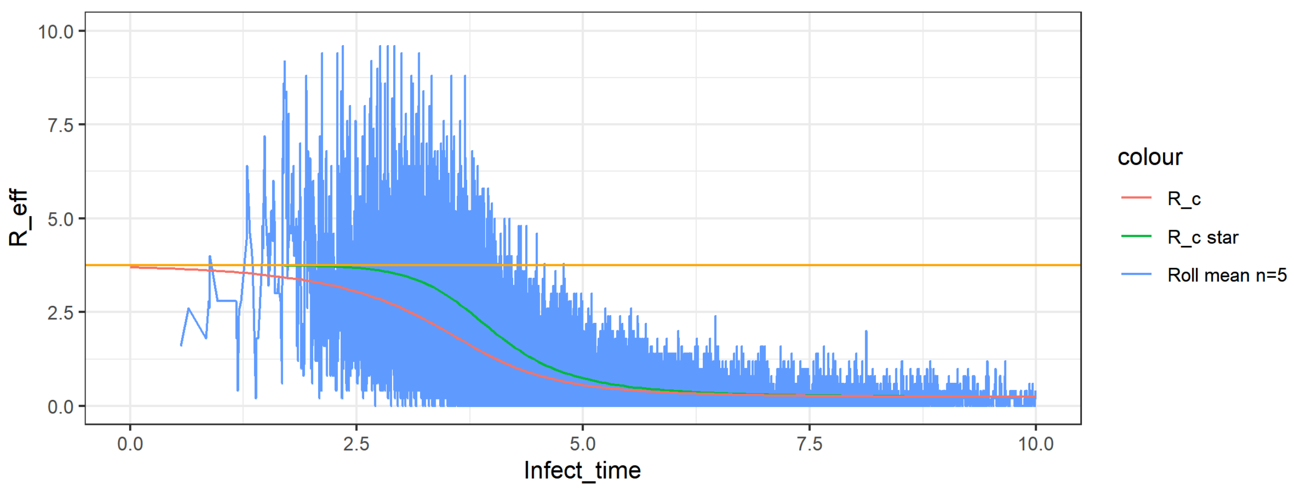
Timely-overlapped 20 simulation



40 simulation:



For larger network size $N = 250,000$, 1 simulation due to time consumption:



TODO: figure out the scale of N s.t. $\mathcal{R}_c(0) \approx \mathcal{R}_c^*(0)$

Thoughts: Comparing ODE for $p(t)$ and $\phi(t)$

$$\frac{d}{dt}p(t) = -\beta\phi_S(t) + (\beta+\gamma)p(t)$$

$$\frac{d}{dt}\phi(t) = +\beta\phi_S(t) - (\beta+\gamma)\phi(t) + \gamma$$

Edge based simulation

JD developed the edge-based Gillespie simulation in R_cpp which seems to be faster for large networks

Negative binomial degree distribution with mean degree $\lambda = 5, \kappa = 0.\ \beta = \gamma = 0.1, N = 10^6$