Network structure induced bias in estimates of intrinsic generation times

Pratyush K. Kollepara[1], Chiara Poletto[2], and Joel C. Miller[1,3]

[1]Department of Mathematical and Physical Sciences, La Trobe University, Melbourne, Australia
[2]Department of Molecular Medicine, University of Padova, Padova, Italy
[3]Australian Centre for AI in Medical Innovation, La Trobe University, Melbourne, Australia

Abstract

The reproduction number is a critical measure of the severity of an infectious disease epidemic. The generation interval, defined as the time taken by an infector to create another infection from its time of infection, is crucial for estimating the basic reproduction number. However, the generation intervals observed, 'realised generation intervals', change over time depending on the dynamics of the epidemic. These time varying distributions are well understood for homogeneous populations, and can be used to infer the intrinsic generation interval distribution. For heterogeneous populations, the state-of-the-art method relies on the use of expensive network based or agent based simulations. We simplify this process by writing exact expressions for simple structured populations. We then use these expressions to demonstrate some previously unexplored biases in the estimation of the intrinsic generation times from the observed generation times, caused by the network structure assumed in the model and particularly the temporal structure of contacts.

1 Introduction

Generation time, the time a parent takes to produce an offspring since its birth, is an important quantity of interest in ecology. The concept has been used extensively in epidemiology, where it refers to the time taken for a newly infected individual to infect someone else. Knowing this interval distribution is necessary for planning contact tracing, quarantine, and isolation in case of an outbreak [13]. It is also necessary to estimate $\mathcal{R}_0$, which quantifies the epidemic transmission potential, from the aggregated incidence of infections using the Wallinga-Lipsitch equation [15]. However, measuring this quantity and correctly interpreting it for outbreak analysis is not straightforward [17].

The Wallinga-Lipstich equation requires the "intrinsic generation interval", which is defined as the time distribution of infectious contacts made by a primary case, or, equivalently, the time distribution of generated infections when the population is fully susceptible and mixes randomly [17]. The observed or realised generation intervals, obtained by contact tracing data, are different from the intrinsic ones, because of dynamical and population mixing effects, and also because of how generation intervals are defined and aggregated from available data. For instance, The forward generation interval distribution refers to the distribution of times it takes for a group of individuals

infected at a given time to transmit the infection to others. On the other hand, the backward generation interval distribution takes the point of view of a cohort of infectees, infected at a given time, and is defined as the distribution of times to infect them [17].

The intrinsic generation interval, determined solely by the infector's infectiousness, remains constant throughout an outbreak. In contrast, forward and backward realized generation times vary. The functional relationship between intrinsic and realised generation time was mathematically studied for a randomly mixing population by the Champredon and Dushoff [17] within the renewal equation framework. The backward generation time increases monotonically, underestimating the intrinsic generation interval early on and overestimating it later. This bias arises from the incidence dynamics. During the growth phase, recent infections dominate the pool of potential infectors, skewing generation times shorter; during the decline, older infections dominate, lengthening the generation intervals. On the other hand, the forward-looking formulation of generation time avoids this dynamic bias. Still, its mean value shrinks right before the peak [17, 16, 12]. One reason for this contraction is the high rate at which susceptibles are being depleted near the peak. An infector may generate multiple transmissions throughout their infectious period and, if the number of susceptibles remains the same, then the transmission events are all equally likely to contribute to the generation intervals [17]. Conversely, if the availability of susceptibles shrinks significantly during the infectious period the later transmissions are less likely to lead to an infection and longer transmission intervals are less likely to be sampled. An alternative mechanism to understand the generation time contraction is the competition between infectors, racing to infect someone [16, 11]. When many infectors compete to infect the same susceptible individual, only the transmission by the fastest infector counts as an infection event and contributes to the generation time statistics, causing the generation interval to contract. The study of these dynamical mechanisms is essential to design techniques to fit the intrinsic generation time from contact tracing data [5, 14, 6].

Although most studies rely on the homogenous mixing assumption, contacts within households, schools, and workplaces are recurrent, heterogeneous, and spatially clustered [19, 3, 9, 10, 7, 8, 1, 2, 19, 4]. Previous studies investigating clustering and recurrence showed that these enhance the competition between infectors and cause a stronger reduction of generation intervals — more than predicted by homogeneous-mixing models — which persists throughout the outbreak [19, 11]. Despite these findings, the effect of the contact network on the generation time dynamics remains under-studied, especially from an analytical perspective. Here, we use the edge-based compartmental modelling framework [18] to derive exact equations for realized generation intervals. The framework allows us to systematically address the role of contacts' heterogeneity - by comparing a homogeneous and a heterogeneous network - and recurrence - by comparing the cases in which contacts' identities continuously change or are fixed. Our theoretical framework, supported by simulations, provides insights into the dynamics of generation intervals and highlights differences between network topologies. Most importantly it points to potential sources of bias, that were left unexplored so far.

---

2 Methods

This section gives a brief overview of the Edge Based Compartmental Modelling (EBCM) framework [18] that is required for understanding the results in the next section. The EBCM framework can be used to solve the SIR model on configuration model networks. These networks are specified by a distribution of the degrees of nodes in the network but are limited to a tree-like structure (do not allow clustering or higher-order structures).

Real-world networks have a temporal nature: some contacts of an individual are fleeting (fast changing) while others are static (long-lasting). The models used here will consider versions of configuration model networks for each type of contact. The first one is known as the annealed or mean field configuration model where even though the number of contacts of an individual remains constant, the edges in the network rewire to new nodes constantly. The other is known as the quenched or static configuration model where the edges are static. In addition to these assumptions about the network structure, we also assume that both transmission and recovery from infection are described by a Poisson process. Detailed derivations of our results are presented in the Supplementary Methods section, and the results can be generalised to non-Poisson processes.

For both annealed and quenched networks, a standard compartmental modelling approach would divide the population up into sub-groups of homogeneous degree (number of contacts), resulting in at least as many differential equations as the number of sub-groups. The EBCM framework reduces such a large system of differential equations into two ordinary differential equations which can be solved numerically with ease.

The EBCM framework makes use of probability-generating functions

$$\psi(x) = \sum_i N_i \; x^i,$$

where N_i refers to the proportion or probability of nodes of degree class i. A variable θ is defined to be the probability that a randomly chosen partnership of a node has not transmitted to it yet. If a node of degree k is susceptible, then none of its k partners have transmitted to it. The probability for the node to be susceptible (given its degree k) is θ^k. The prevalence of susceptibles in the population can be computed using

$$S = \sum_k N_k \theta^k = \psi(\theta).$$

2.1 Annealed network

Homogeneous compartmental models use macroscopic variables such as the proportion of the population that is infected, susceptible, recovered etc. The probability a random interaction with another individual is with an infected individual is assumed to equal the proportion of the population that is infected. In a more general case where the number of partners differs among individuals, the probability a random contact is with an infected individual is the prevalence of half-edges in the population that are attached to an infected node. The variable $I(t)$ (the proportion of the population that is infected) is generalised to a variable $\pi_I(t)$ which is the proportion of half-edges that connect to infected nodes. Similarly, we generalize $\pi_S(t)$ and $\pi_R(t)$ which are the prevalence of half-edges with susceptible and recovered nodes respectively. We also define a new variable $\pi_i(t)$, which is the incidence of infected stubs, i.e. the instanteous rate at which half-edges with infected nodes are created by new infections.

We find that the probability a random half-edge connects to a susceptible node is

$$\pi_S = \frac{\sum_k kN_k\theta^k}{\sum_k kN_k} = \theta\frac{\psi'(\theta)}{\psi'(1)}$$

Knowing π_S requires $\theta(t)$. At the start of the epidemic, the initially susceptible nodes have not received any transmissions. So, using the initial condition $\theta(0) \approx 1$, the following differential equation (derived in ref [18]) can be solved to find $\theta(t)$:

$$\dot{\theta} = -\beta\theta + \beta\frac{\theta^2\psi'(\theta)}{\psi'(1)} - \gamma\theta\ln\theta$$

2.2 Quenched network

In the quenched case, the variable $S(t)$ is generalised to a variable $\phi_S(t)$ which is the probability that the partner of a randomly selected node is susceptible (under the assumption that the randomly selected node has not transmitted to its partner).

$$\phi_S = \frac{\sum_k kN_k\theta^{k-1}}{\sum_k kN_k} = \frac{\psi'(\theta)}{\psi'(1)}$$

In this case, using a similar initial condition for θ as before, the following differential equation [18] is solved:

$$\dot{\theta} = -\beta\theta + \beta\frac{\psi'(\theta)}{\psi'(1)} - \gamma(1-\theta)$$

---

3 Results

3.1 Annealed network

In a homogeneous population, the forward generation interval distribution can be calculated using the intrinsic generation interval distribution, $g(\tau)$, and the proportion or number of susceptible people, $S(t)$, with time [17].

$$f_H(t,\tau) = \frac{g(\tau)S(t+\tau)}{\int_0^\infty dx\ g(x)S(t+x)}$$

Here, $S(t)$ gives the probability for a random individual to come in contact with a susceptible individual at time t. In an annealed heterogeneous network this is no longer $S(t)$ and depends on the connectivity. We thus need to replace it with $\pi_S(t)$. Therefore, the forward generation interval distribution for an annealed configuration model network is

$$f_A(t,\tau) = \frac{g(\tau)\pi_S(t+\tau)}{\int_0^\infty dx\ g(x)\pi_S(t+x)}.$$

Turning our attention to the backward generation interval, Champredon and Dushoff's equation for the homogeneous case makes use of the incidence of new infections as a function of time [17]:

$$b_H(t,\tau) = \frac{g(\tau)i(t-\tau)}{\int_0^\infty dx\ g(x)i(t-x)}.$$

If we use the average incidence of stubs, instead of the incidence of infections, we obtain an exact expression for the backward generation interval distribution for an annealed heterogeneous network.

$$b_A(t,\tau) = \frac{g(\tau)\pi_i(t-\tau)}{\int_0^\infty dx\ g(x)\pi_i(t-x)}.$$

A general characteristic of the backward generation intervals is that they are stable in time when the incidence grows exponentially [17]. This remains valid in the heterogeneous network case, as can be seen by plugging in $\pi_i(t) = \pi_i(0)e^{\lambda t}$ into the above formula, leading to

$$b_A^{exp}(t,\tau) = (\lambda+\gamma)e^{-(\lambda+\gamma)\tau}$$

Similar to the forward generation interval, we define an effective reproduction number called the forward reproduction number, $\mathcal{R}_f$. It represents the expected number of infections created by an average infector who became infected at time t over its entire infectious period. For an annealed network,

$$\mathcal{R}_f^A(t) = \frac{\beta(\theta\psi''(\theta(t))+\psi'(\theta(t)))}{\psi'(\theta(t))} \int_0^\infty d\tau \; e^{-\gamma\tau}\pi_S(t+\tau)$$

If a reproduction number is calculated empirically through observation of transmission chains, we would expect it to match with the forward reproduction number.

3.2 Quenched network

In an annealed network, a node continuously breaks its edges and re-connects to new edges. So, if a node infects another node, it does not compete with itself when transmitting again. Conversely, in a quenched network, a node can compete with itself — after the first transmission all later transmissions to the same neighbour have no effect. In the equations for the realised generation time we thus need to replace the intrinsic generation time $g(\tau)$ with $\tilde{G}(\tau) = e^{-(\beta+\gamma)\tau}$. In addition, the probability that a stub points to a susceptible node – which was $\pi_S(t)$ in the annealed case – is now given by $\phi_S(t)$. The modified version of the equation in [17] then becomes

$$f_Q(t,\tau) = \frac{\tilde{G}(\tau)\phi_S(t+\tau)}{\int_0^\infty dx \; \tilde{G}(x)\phi_S(t+x)}.$$

In the backward case, the incidence of infections, $i(t)$ is replaced with the average probability of incidence among the neighbours of a node, $\phi_i(t)$:

$$b_Q(t,\tau) = \frac{\tilde{G}(\tau)\phi_i(t-\tau)}{\int_0^\infty dx \; \tilde{G}(x)\phi_i(t-x)}$$

From this equation, we obtain that the backward generation intervals are stable when the epidemic grows exponentially, similar to the annealed case. The distribution in this regime is

$$b_Q^{exp}(t,\tau) = (\lambda+\beta+\gamma)e^{-(\lambda+\beta+\gamma)\tau}$$

Finally, the forward reproduction number can be derived for quenched network as follows

$$\mathcal{R}_f^Q(t) = \frac{\beta\psi''(\theta(t))}{\psi'(\theta(t))} \int_0^\infty d\tau \; e^{-(\beta+\gamma)\tau}\phi_S(t+\tau)$$

3.3 Impact of network properties on the realised generation time

We validate the exact results using event-based simulations [21]. Figures 1 and 2 show the comparison between analytical solutions and simulations for the annealed and quenched case, respectively. In each figure, we plot the epidemic profile, the forward and backward generation times and the forward reproduction number. The figures show the comparison for two different topologies that we call TMD and TPL70. TMD refers to a tri-modal degree distribution and TPL70 refers to a truncated power law degree distribution (See Figure captions more details). Both networks are heterogeneous, but the TPL70 network is a case of extreme heterogeneity. The forward generation intervals in the annealed case show a contraction qualitatively similar to what is predicted by the homogenous mixing equations [17]. This is, however, less evident when the network is extremely heterogeneous. In the quenched case forward generation times shrink substantially due to the intensified competition

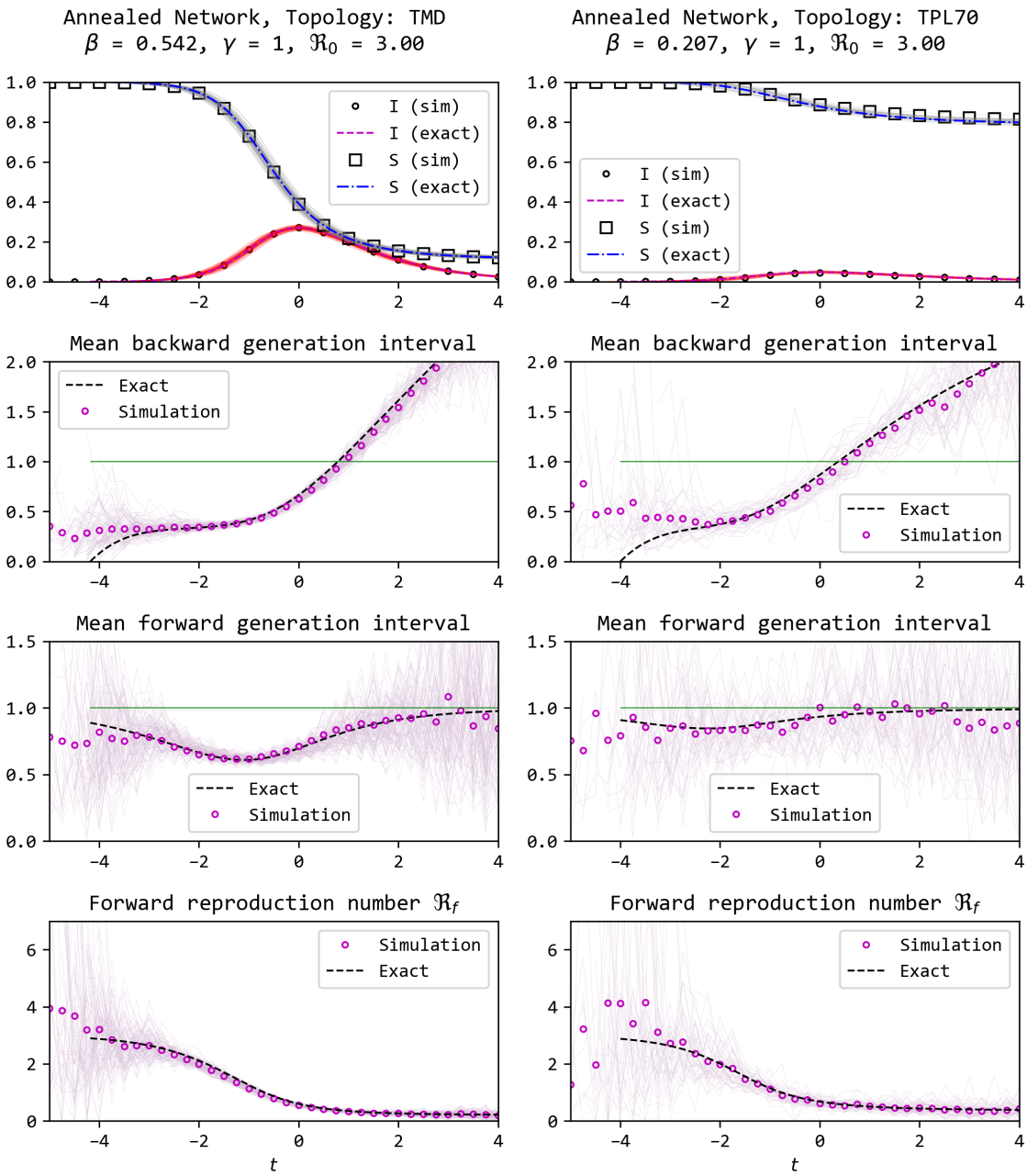

Figure 1: *Comparison of epidemics and their realised generation intervals using an event-based simulation method and the exact solution for annealed networks with two types of heterogeneous degree distributions (Left: the degree of a node can be 3, 5, or 7 with equal probabilities, Right: degree distribution is a truncated power law with maximum degree of 70 and exponent of -2). We find a good match between theory and simulations for the realised generation intervals and the forward reproduction number. The time-series are centred such that the prevalence peaks at time zero. The thin lines show trajectories from each of the 100 simulations, while the plot with markers shows the average. At the start and towards the end of the epidemic, significant noise is observed and the simulation deviates from the theory.*

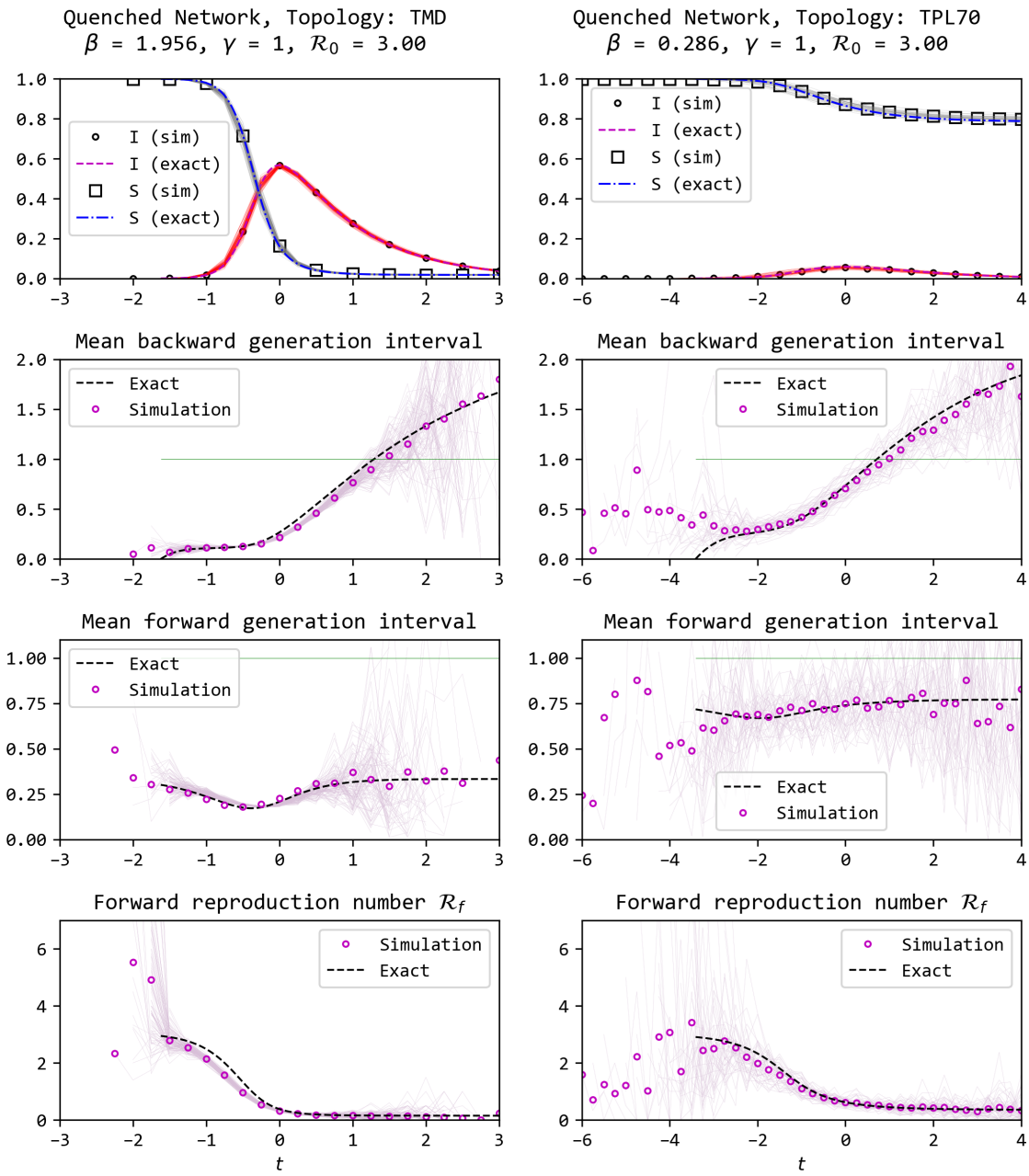

Figure 2: *Comparison of epidemics and their realised generation intervals using an event-based simulation method and the exact solution for quenched networks with two types of heterogeneous degree distributions (Left: the degree of a node can be 3, 5, or 7 with equal probabilities, Right: degree distribution is a truncated power law with maximum degree of 70 and exponent of -2). We find a good match between theory and simulations for the realised generation intervals and the forward reproduction number. The time-series are centred such that the prevalence peaks at time zero. The thin lines show trajectories from each of the 100 simulations, while the plot with markers shows the average. At the start and towards the end of the epidemic, significant noise is observed and the simulation deviates from the theory for the truncated power law distribution.*

among infectors. While forward generation time varies substantially from one network to another, backwards generation time is quite robust to the network structure.

We can further analyse the impact of the network features by numerically solving the derived analytical expressions for different scenarios. This is significantly inexpensive compared to stochastic numerical simulations. We performed various numerical experiments and we summarise the results here:

Experiment (i) Structure: Annealed TPL70 network and homogeneous network with same mean degree. Disease parameters: Identical recovery rate, intrinsic generation interval ($\gamma = 1$) and $\mathcal{R}_0$. Result: Heterogeneity reduces the contraction in the forward generation interval. However, this could possibly be explained by the lower peak (hence lesser competition) in the heterogeneity (See supplementary results).

Experiment (ii) Structure: Annealed TPL70 network and homogeneous network with same mean degree.
Disease parameters: Identical recovery rate, intrinsic generation interval ($\gamma = 1$). The transmission rate is selected such that the two epidemics have the same level of peak incidence of infections.
Result: Heterogeneity increases the contraction in the forward generation interval, despite the same maximum incidence of infections. This can be explained by noting that the peak of $\pi_i(t)$ is higher than $i(t)$. Furthermore, the backward generation intervals also differ significantly (See supplementary results).

Experiment (iii) Structure: Annealed TPL70 network Method: Compare distribution of realised intervals assuming homogeneity [17] (i.e. using $i(t)$ and $S(t)$) and not assuming homogeneity (i.e. using $\pi_i(t)$ and $\pi_S(t)$)
Result: Using the net incidence of infections underestimates the contraction in forward generation interval. In contrast, the backward generation intervals are relatively unchanged (See supplementary results).

3.4 Implications for estimation of reproduction number

In this section, we will explore how errors could arise if disease parameters are inferred using existing methods to relate the observed generation interval and growth rate to the basic reproduction number.

If the intrinsic generation intervals and the exponential growth rate are known, the basic reproduction number can be estimated using the Wallinga-Lipsitch equation [15]. The growth rate can be estimated from incidence at the start of the epidemic. However, the intrinsic generation intervals are not observed. Instead we observe the realised generation intervals (forward or backward) which are influenced by the dynamics of the outbreak and the contact structure of the population.

So to fit model parameters and infer the basic reproduction number from the growth rate and the observed generation intervals, it is crucial to recognise the distinction between the realised generation intervals and the intrinsic generation intervals. In principle, the exact expressions for realised generation interval distributions can be used to infer the intrinsic generation interval distribution, which can be used to then estimate $\mathcal{R}_0$ [17, 5, 14, 6]. However, the Champredon and Dushoff's equation and further developments relies on the homogenous mixing assumption.

Our results from previous sectionss show that for a heterogeneous population, the conclusions will be incorrect if simple population aggregates such as prevalence of susceptibles or incidence of

infections are used. We expect that the joint use of simple aggregate observations of an outbreak and realised generation intervals could bias the estimation of the basic reproduction number and the intrinsic generation intervals. To demonstrate this bias, we now look at an example case of outbreak analysis and two scenarios with different amounts of information available to the disease modeller.

Example Let us consider a pathogen causing an epidemic in different disconnected populations. There is a population in which the authorities have the capacity and resources to document the incidence of cases and also perform backward contact tracing (let us call it Population 1). There is another population where we do not have the capacity to perform contact tracing, but we can document case incidence (call it Population 2). Both populations are well described by a quenched and heterogeneous contact network. This is a simplification, as any real-world network will have a mixture of short and long time scale contacts. First, we describe the quantities of interest in the ground reality from the correct model. Then, we consider cases where an infectious disease modeller has some information about the contact structure and see how their modelling assumptions can affect model inference in both populations.

Using the ground reality model, we can write the basic reproduction numbers in terms of the observable quantities:

$$\mathcal{R}_0^1 = \frac{1}{1 - \lambda_1 T_{exp,1}} \tag{1}$$

$$\mathcal{R}_0^2 = \frac{1}{1 - \lambda_2 T_{exp,2}} \tag{2}$$

where the λ is the exponential growth rate estimated from incidence and T_{exp} is the mean backward generation interval during the exponential growth phase, estimated from the contact tracing data. We can also write the average intrinsic generation time as a function of the same quantities. However, the expression depends also on the first and second moments of the degree distribution. See detailed methods for a detailed calculation.

Case A: The modeller does not know anything about the contact structure and assumes a well-mixed homogeneous model. The model has essentially two parameters, the transmission rate β and the recovery rate γ (which also specifies the intrinsic generation intervals). Using this model, the estimators for population 1 are:

$$\widehat{\gamma}_1 = \frac{1}{T_{exp,1}} - \lambda_1 \tag{3}$$

$$\widehat{\mathcal{R}_0^1} = \frac{1}{1 - \lambda_1 T_{exp,1}} \tag{4}$$

Comparing these estimates to the ground reality tells us that the basic reproduction number is unbiased but $\widehat{\gamma}_1$ is not as it does not properly account for the contact network properties – first and second moment of the degree distribution.

For population 2, nothing is known about the backward generation intervals, but the modeller has an estimate for the intrinsic generation intervals in population 1. Since by definition, intrinsic intervals would not vary among populations, the modeller might see it fit to treat $\widehat{\gamma}_1$ as an estimator for γ_2. This leads to an estimate of the reproduction number

$$\widehat{\mathcal{R}_0^2} = \frac{1 + (\lambda_2 - \lambda_1) T_{exp,1}}{1 - \lambda_1 T_{exp,1}} \neq \frac{1}{1 - \lambda_2 T_{exp,2}} \tag{5}$$

which is biased. We find that the bias in the estimator for population 2 is greater when the contact structure is different among the two populations. When the contact structure is identical, the estimator is unbiased when the growth rates are equal.

Case B: The modeller knows that the contacts are quenched and also has correct information of the average number of contacts $\langle K_1 \rangle$ and $\langle K_2 \rangle$ respectively. In the absence of more detailed information, they assume a homogeneous quenched network with degrees $k_1 = \langle K_1 \rangle$ and $k_2 = \langle K_2 \rangle$, respectively. The model has two other parameters, the transmission rate β and the recovery rate γ (which also specifies the intrinsic generation intervals). Using this model, the estimators for population 1 are:

$$\widehat{\gamma}_1 = \frac{(k_1 - 2)(1/T_{exp} - \lambda) - \lambda}{k_1 - 1} \tag{6}$$

$$\widehat{\mathcal{R}_0^1} = \frac{1}{1 - \lambda_1 T_{exp,1}} \tag{7}$$

Comparison with the ground reality model shows that $\widehat{\gamma}_1$ is biased, as calculation does not correctly account for the second moment of the degree distribution, while $\widehat{\mathcal{R}_0^1}$ is not. As in the previous case, the modeller might see fit to treat $\widehat{\gamma}_1$ as an estimator for γ_2. This leads to an estimate of the reproduction number for population 2

$$\widehat{\mathcal{R}_0^2} = \frac{(\lambda_2 + \widehat{\gamma}_1)(k_2 - 1)}{\lambda_2 + \widehat{\gamma}_1(k_2 - 1)}$$

This estimator is biased, albeit behaves very differently than the biased estimator in Case A.

In summary, we find that if the incidence and backward contact tracing is available in a population, then even incorrect assumptions about the contact structure can yield an unbiased estimator for the basic reproduction number, even though the estimator for the intrinsic generation interval is biased. Using this estimated intrinsic generation interval in order to estimate the basic reproduction number in another population leads to biased estimates. See 3 for an illustration of the bias in $\widehat{\mathcal{R}_0^2}$.

While case A is realistic, case B is hypothetical since more information – i.e. the average number of contacts – is available and used in the analysis than what normally happens in reality. Paradoxically, case B presents a more severe bias. For Population 1, a truncated power law degree distribution is used with an exponent of -2 and maximum degree of 70. Population 2 has a similar structure but with a maximum degree of 30. The growth rate in Population 1 is $\lambda_1 = 0.5$ and recovery rate is $\gamma = 1$.

---

4 Discussion

In this article, we have generalised Champredon and Dushoff's theory [17] for realised generation intervals in a homogeneous and well-mixed population to heterogeneous networks, using the EBCM framework. This framework allows incorporating degree heterogeneity in a tree-like network. In structured populations, realised generation intervals have been investigated using network and agent-based simulations [19, 4]. These methods benefit from their ability to be modified quite easily – adding more coarse-grained structure or even including forms of structure beyond degree heterogeneity. However, they are slow and managing the noise reasonably would require a large number of simulations or a large population. In addition, they do not provide analytical insights. Some

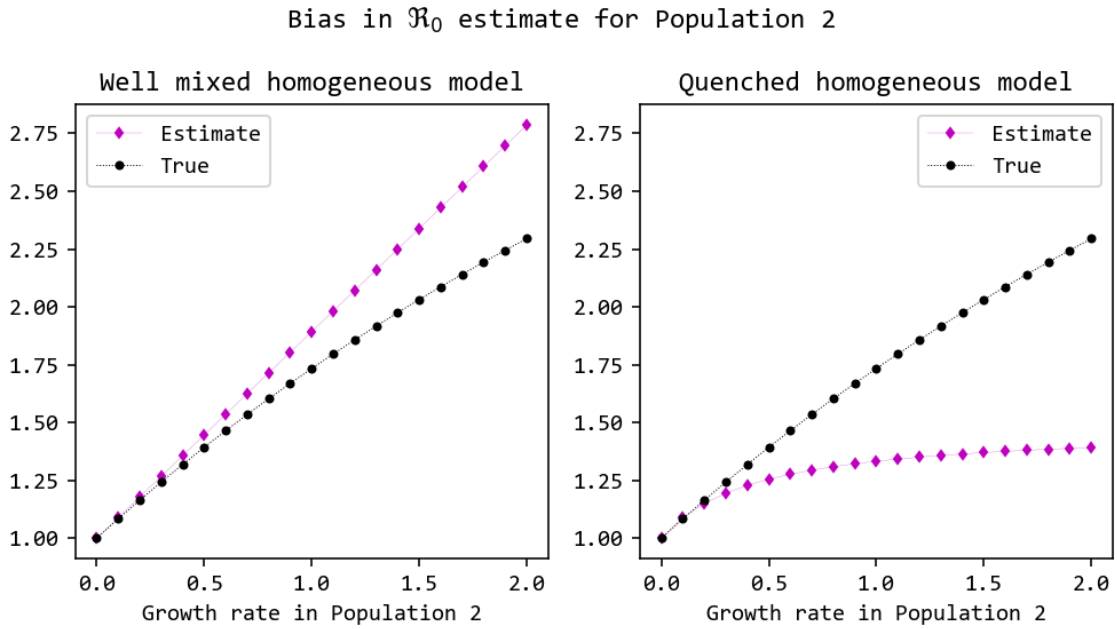

Figure 3: *Comparison between the estimated and true value of the basic reproduction number in a heterogeneous quenched network when different homogeneous models are used and the intrinsic generation interval is estimated from a different population (see 3.4 for a detailed explanation).*

analytical considerations were provided by a few works [20, 11]. Our exact expressions, validated using simulations, present an advance in our ability to predict generation intervals in simple heterogeneous contact structures and sets up a foundation to obtain exact expressions for more complex heterogeneous networks.

Generation intervals are useful in estimating the basic reproduction number of an epidemic. Champredon and Dushoff showed how fitting the mean intrinsic generation interval to the mean of the realised generation intervals (backward) is incorrect [17]. They anticipate that heterogeneity may affect their results. We confirm this, determine the manner by which it affects the results and more importantly, through what mechanism. We found that the realised intervals depend not on simple aggregate measures like net prevalence of susceptibles or net incidence of infection, but on more involved measures such as the prevalence of half-edges in the network which have a susceptible or a newly infected node attached to them. Measurement of such quantities requires higher resolution observation of the population, at the scale of groups which are homogeneous in number of contacts.

We also use a series of hypothetical experiments to show that the estimates of intrinsic intervals (inferred using realised generation intervals) are sensitive to model assumptions. Despite this sensitivity, the estimates of $\mathcal{R}_0$ can be relied upon as long as the estimated intrinsic generation intervals are used in the same setting where they were estimated. If the estimated intrinsic intervals are used in a different setting (say an epidemic caused by the same pathogen in a different setting where the contact/network structure differs) then the estimates of $\mathcal{R}_0$ will be biased. This practice of applying generation interval estimates or related parameter estimates from from one setting to an epidemic of the same pathogen in another setting is common [22, 23, 24, 25, 26, 27].

The forward reproduction number that was derived here is an effective reproduction number. Effective reproduction numbers are of two types: instantaneous or cohort-based (case reproduction number). The instantaneous reproduction number measures the transmission happening at certain time while the case reproduction number measures the transmission that is caused by a cohort of infectors who became infected at a certain time [28]. The forward reproduction number, $\mathcal{R}_f$, is equivalent to the case reproduction number [28, 29].

A key assumption throughout this work is that the recovery from infection is a Poisson process (the intrinsic generation interval distribution is exponential). While real-world diseases do not follow this assumption, the expressions for forward and backward generation intervals that we derived are fairly general. Real-world networks have many kinds of heterogeneities beyond the kind we have investigated here. They have clustering (triads that form a triangle), multiple layers and temporal variation. Here, we have only studied degree heterogeneity in two extreme cases of temporal variation – quenched (static partners) and annealed (fleeting contact with changing partners). In principle, a linear combination of the distributions derived here may be used to model contact networks with a more realistic temporal behaviour.

Data and Software

The software used in this study is available at
`https://github.com/praty-k/gi-networks`

Acknowledgements

We thank Andrew Black and Eben Kenah for helpful conversations. This research was supported by use of the Nectar Research Cloud, a collaborative Australian research platform supported by the NCRIS-funded Australian Research Data Commons (ARDC). P.K.K. was supported by the La Trobe Graduate Research Scholarship and Full-Fee Research Scholarship and the Ed Smith Applied Mathematics Fellowship. C.P. was supported by the Cariparo Foundation through the program Starting Package and the Department of Molecular Medicine through the program SID from BIRD funding.

References

[1] Barrat, A., Cattuto, C., Colizza, V., Gesualdo, F., Isella, L., Pandolfi, E., Pinton, J.-F., Ravà, L., Rizzo, C., Romano, M., Stehlé, J., Tozzi, A. E., & Van den Broeck, W. (2013). Empirical temporal networks of face-to-face human interactions. *The European Physical Journal Special Topics*, 222(6), 1295–1309. https://doi.org/10.1140/epjst/e2013-01927-7

[2] Eames, K., Bansal, S., Frost, S., & Riley, S. (2015). Six challenges in measuring contact networks for use in modelling. *Epidemics*, 10, 72–77. https://doi.org/10.1016/j.epidem.2014.08.006

[3] Aleta, A., Martín-Corral, D., Bakker, M. A., Pastore y Piontti, A., Ajelli, M., Litvinova, M., Chinazzi, M., Dean, N. E., Halloran, M. E., Longini, I. M., Pentland, A., Vespignani, A., Moreno, Y., & Moro, E. (2022). Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas. *Proceedings of the National Academy of Sciences*, 119(26), e2112182119. https://doi.org/10.1073/pnas.2112182119

[4] Pung, R., Russell, T. W., & Kucharski, A. J. (2024). Detecting changes in generation and serial intervals under varying pathogen biology, contact patterns and outbreak response. *PLOS Computational Biology*, 20(3), e1011967. https://doi.org/10.1371/journal.pcbi.1011967

[5] Park, S. W., Sun, K., Abbott, S., Sender, R., Bar-on, Y. M., Weitz, J. S., Funk, S., Grenfell, B. T., Backer, J. A., Wallinga, J., Viboud, C., & Dushoff, J. (2023). Inferring the differences in incubation-period and generation-interval distributions of the Delta and Omicron variants of SARS-CoV-2. *Proceedings of the National Academy of Sciences*, 120(22), e2221887120. https://doi.org/10.1073/pnas.2221887120

[6] Chen, D., Lau, Y.-C., Xu, X.-K., Wang, L., Du, Z., Tsang, T. K., Wu, P., Lau, E. H. Y., Wallinga, J., Cowling, B. J., & Ali, S. T. (2022). Inferring time-varying generation time, serial interval, and incubation period distributions for COVID-19. *Nature Communications*, 13(1), 7727. https://doi.org/10.1038/s41467-022-35496-8

[7] Loedy, N., Wallinga, J., Hens, N., & Torneri, A. (2024). Repetition in social contacts: Implications in modelling the transmission of respiratory infectious diseases in pre-pandemic and pandemic settings. Proceedings of the Royal Society B: Biological Sciences, 291(2027), 20241296. https://doi.org/10.1098/rspb.2024.1296

[8] Pung, R., Firth, J. A., Russell, T. W., Rogers, T., Lee, V. J., & Kucharski, A. J. (2024). Temporal contact patterns and the implications for predicting superspreaders and planning of targeted outbreak control. Journal of The Royal Society Interface, 21(221), 20240358. https://doi.org/10.1098/rsif.2024.0358

[9] Moreno López, J. A., Arregui García, B., Bentkowski, P., Bioglio, L., Pinotti, F., Boëlle, P.-Y., Barrat, A., Colizza, V., & Poletto, C. (2021). Anatomy of digital contact tracing: Role of age, transmission setting, adoption and case detection. Science Advances, 7, eabd8750. https://doi.org/10.1126/sciadv.abd8750

[10] Calmon, L., Colosi, E., Bassignana, G., Barrat, A., & Colizza, V. (2024). Preserving friendships in school contacts: An algorithm to construct synthetic temporal networks for epidemic modelling. PLOS Computational Biology, 20(12), e1012661. https://doi.org/10.1371/journal.pcbi.1012661

[11] Park, S. W., Champredon, D., & Dushoff, J. (2020). Inferring generation-interval distributions from contact-tracing data. Journal of The Royal Society Interface, 17(167), 20190719. https://doi.org/10.1098/rsif.2019.0719

[12] Svensson, Å. (2007). A note on generation times in epidemic models. Mathematical Biosciences, 208(1), 300–311. https://doi.org/10.1016/j.mbs.2006.10.010

[13] Kraemer, M. U. G., Pybus, O. G., Fraser, C., Cauchemez, S., Rambaut, A., & Cowling, B. J. (2021). Monitoring key epidemiological parameters of SARS-CoV-2 transmission. Nature Medicine, 27(11), 1854–1855. https://doi.org/10.1038/s41591-021-01545-w

[14] Manica, M., De Bellis, A., Guzzetta, G., Mancuso, P., Vicentini, M., Venturelli, F., Zerbini, A., Bisaccia, E., Litvinova, M., Menegale, F., Grané, C. M., Poletti, P., Marziano, V., Zardini, A., d'Andrea, V., Trentini, F., Bella, A., Riccardo, F., Pezzotti, P., Ajelli, M., Rossi, P. G., & Merler, S. (2022). Intrinsic generation time of the SARS-CoV-2 Omicron variant: An observational study of household transmission. The Lancet Regional Health – Europe, 19, 100446. https://doi.org/10.1016/j.lanepe.2022.100446

[15] Wallinga, J., & Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. Proceedings of the Royal Society B: Biological Sciences, 274(1609), 599–604. https://doi.org/10.1098/rspb.2006.3754

[16] Kenah, E., Lipsitch, M., & Robins, J. M. (2008). Generation interval contraction and epidemic data analysis. Mathematical Biosciences, 213(1), 71–79. https://doi.org/10.1016/j.mbs.2008.02.007

[17] Champredon, D., & Dushoff, J. (2015). Intrinsic and realized generation intervals in infectious-disease transmission. Proceedings of the Royal Society B: Biological Sciences, 282(1821), 20152026. https://doi.org/10.1098/rspb.2015.2026

[18] Miller, J. C., Slim, A. C., & Volz, E. M. (2011). Edge-based compartmental modelling for infectious disease spread. Journal of The Royal Society Interface, 9(70), 890–906. https://doi.org/10.1098/rsif.2011.0403

[19] Liu, Q.-H., Ajelli, M., Aleta, A., Merler, S., Moreno, Y., & Vespignani, A. (2018). Measurability of the epidemic reproduction number in data-driven contact networks. Proceedings of the National Academy of Sciences, 115(50), 12680–12685. https://doi.org/10.1073/pnas.1811115115

[20] Trapman, P., Ball, F., Dhersin, J.-S., Tran, V. C., Wallinga, J., & Britton, T. (2016). Inferring R 0 in emerging epidemics—The effect of common population structure is small. Journal of The Royal Society Interface, 13(121), 20160288. https://doi.org/10.1098/rsif.2016.0288

[21] Kiss, I. Z., Miller, J. C., & Simon, P. L. (2017). Mathematics of Epidemics on Networks: From Exact to Approximate Models (Vol. 46). Springer International Publishing. https://doi.org/10.1007/978-3-319-50806-1

[22] Saldaña, F., Daza-Torres, M. L., & Aguiar, M. (2023). Data-driven estimation of the instantaneous reproduction number and growth rates for the 2022 monkeypox outbreak in Europe. PLOS ONE, 18(9), e0290387. https://doi.org/10.1371/journal.pone.0290387

[23] Díaz-Brochero, C., & Cucunubá, Z. M. (2024). Epidemiological findings, estimates of the instantaneous reproduction number, and control strategies of the first Mpox

outbreak in Latin America. Travel Medicine and Infectious Disease, 59, 102701. https://doi.org/10.1016/j.tmaid.2024.102701

[24] Shim, E., Tariq, A., Choi, W., Lee, Y., & Chowell, G. (2020). Transmission potential and severity of COVID-19 in South Korea. International Journal of Infectious Diseases, 93, 339–344. https://doi.org/10.1016/j.ijid.2020.03.031

[25] Rocklöv, J., Sjödin, H., & Wilder-Smith, A. (2020). COVID-19 outbreak on the Diamond Princess cruise ship: Estimating the epidemic potential and effectiveness of public health countermeasures. Journal of Travel Medicine, 27(3), taaa030. https://doi.org/10.1093/jtm/taaa030

[26] Munayco, C. V., Tariq, A., Rothenberg, R., Soto-Cabezas, G. G., Reyes, M. F., Valle, A., Rojas-Mezarina, L., Cabezas, C., Loayza, M., Chowell, G., Garro, D. C., Vasquez, K. M., Castro, E. S., Ordinola, I. S., Mimbela, J. M., Cornejo, K. M., Quijano, F. C., La Torre Rosillo, L., Ibarguen, L. O., ... Ramos, W. (2020). Early transmission dynamics of COVID-19 in a southern hemisphere setting: Lima-Peru: February 29th–March 30th, 2020. Infectious Disease Modelling, 5, 338–345. https://doi.org/10.1016/j.idm.2020.05.001

[27] Salvatore, M., Basu, D., Ray, D., Kleinsasser, M., Purkayastha, S., Bhattacharyya, R., & Mukherjee, B. (2020). Comprehensive public health evaluation of lockdown as a non-pharmaceutical intervention on COVID-19 spread in India: National trends masking state-level variations. BMJ Open, 10(12), e041778. https://doi.org/10.1136/bmjopen-2020-041778

[28] Gostic, K. M., McGough, L., Baskerville, E. B., Abbott, S., Joshi, K., Tedijanto, C., Kahn, R., Niehus, R., Hay, J. A., De Salazar, P. M., Hellewell, J., Meakin, S., Munday, J. D., Bosse, N. I., Sherrat, K., Thompson, R. N., White, L. F., Huisman, J. S., Scire, J., ... Cobey, S. (2020). Practical considerations for measuring the effective reproductive number, Rt. PLOS Computational Biology, 16(12), e1008409. https://doi.org/10.1371/journal.pcbi.1008409

[29] Fraser, C. (2007). Estimating Individual and Household Reproduction Numbers in an Emerging Epidemic. PLoS ONE, 2(8), e758. https://doi.org/10.1371/journal.pone.0000758