Journal of Theoretical Biology 618 (2026) 112314

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/jtbi

Reducing size bias in epidemic network modelling

Neha Bansal *, Katerina Kaouri, Thomas E. Woolley

School of Mathematics, Cardiff University, Senghennydd Road, Cardiff, CF24 4AG, UK

ARTICLE INFO

Keywords:
Networks
SIR Model
Sampling
Disease modelling
Size bias
Cattle network
Policymaking
Interventions
Human contact networks

ABSTRACT

Epidemiological models can inform policymaking on disease control strategies, and these models often rely on sampled contact networks. The Random Walk (RW) sampling algorithm, commonly used for network sampling, produces size-biased samples that over-represent highly connected individuals, leading to biased estimates of disease spread. The Metropolis-Hastings Random Walk (MHRW) addresses this by providing samples representative of the underlying network's connectivity distribution. We compare MHRW and RW in reducing size bias across four network types: Erdös-Rényi (ER), Small-world (SW), Negative-binomial (NB), and Scale-free (SF). We simulate disease spread using a stochastic Susceptible-Infected-Recovered (SIR) framework. RW tends to overestimate infections (by 25 % in ER, SW, NB) and secondary infections (by 25 % in ER, SW and 80 % in NB), and underestimate time-to-infection in NB networks. MHRW reduces the size bias, except on SF networks, where both algorithms provide non-representative samples and highly variable estimates. We find that RW is appropriate for fast-spreading, high-mortality epidemics in homogeneous or moderately random networks (ER, SW). In contrast, MHRW is better suited for slower and low-severity epidemics and can be effective in both homogeneous and heterogeneous networks (ER, SW, NB). However, MHRW is computationally expensive and less accurate when duplicate nodes are removed. We also analyse real-world data from cattle movement and human contact networks; MHRW generates disease spread estimates closer to the underlying network than RW. Our findings guide the selection of sampling algorithms based on network structure and epidemic characteristics, enhancing the reliability of disease modelling for policymaking.

1. Introduction

During an epidemic, accurately estimating metrics of disease spread, such as the number of infected individuals (epidemic size), effective reproduction number (secondary infections), and time-to-infection, is crucial for effective disease mitigation (Nunes et al., 2024). Epidemic modelling provides a suite of tools, frameworks, and methodologies to generate or predict disease metric estimates (Vynnycky and White, 2010). However, estimating these metrics on real-world networks poses challenges due to the significant costs and time required to gather data for an entire population, along with the high computational cost of processing large datasets. Typically, a sample from the real-world network is used for epidemic modelling. But, inconsistencies between the sampling method and the underlying contact network may result in biased and non-representative samples, leading to skewed estimates of disease metrics (Suhail et al., 2021; Joyal-Desmarais et al., 2022).

Traversal-based sampling (TBS) algorithms are widely used for sampling networks (Gjoka et al., 2010; Craft and Caillaud, 2011; Malmros et al., 2016; Fournet and Barrat, 2017; Cui et al., 2022). These methods begin with one or more initial nodes (called seeds) and choose the next node to sample based on specific information about their neighbours. The primary distinction among TBS methods lies in selecting the next node.

One notable TBS algorithm, the Random Walk (RW) algorithm (Wei et al., 2004), is often used in epidemiological studies of large populations (Milligan et al., 2004; Qi et al., 2023) due to its low computational cost. Specifically, for infectious diseases, the RW algorithm works well for homogeneous or moderately random contact networks, where individuals have a similar number of contacts or degree on average. However, real-world contact networks are often heterogeneous (Nielsen et al., 2020), meaning that individuals exhibit varying levels of heterogeneity. This heterogeneity can result in size-biased samples (Arratia et al., 2019) of the underlying network (UN) when using the RW algorithm (Stein et al., 2014; Li et al., 2015), leading to the over-representation of highly connected individuals. This size bias in samples can lead to inaccurate disease spread estimates, especially when the risk of infection correlates with an individual's degree, as for COVID-19, measles, chickenpox, etc., where the number of contacts is a critical factor in the spread of the virus.

The Metropolis-Hastings Random Walk (MHRW) (Hu and Lau, 2013) algorithm reduces size bias by adjusting the selection probability of each individual based on its degree, thereby reducing the likelihood of

* Corresponding author.
E-mail addresses: bansaln3@cardiff.ac.uk (N. Bansal), KaouriK@cardiff.ac.uk (K. Kaouri), WoolleyT1@cardiff.ac.uk (T.E. Woolley).

https://doi.org/10.1016/j.jtbi.2025.112314
Received 29 January 2025; Received in revised form 22 October 2025; Accepted 23 October 2025
Available online 15 November 2025

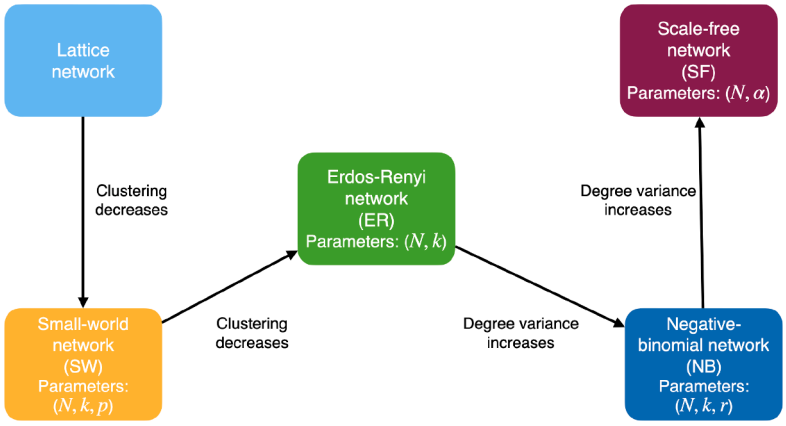


Fig. 1. A simple flow of change in network types, illustrating the transition from ER to SW and Lattice networks with increasing clustering, and from ER to NB and SF networks with the increase in variance of the degree distribution or degree heterogeneity. Network size is denoted by N, average degree by k, edge creation probability by p, dispersion parameter by r, and power-law exponent by α, see Appendix A.1 for a detailed description.

sampling highly connected individuals. This adjustment allows MHRW to generate samples that more accurately reflect the degree distribution of the UN. However, MHRW is approximately $1.5 - 2$ times more computationally expensive than RW (Bishop and Nasrabadi, 2006).

Here, we compare the effectiveness of MHRW and RW in reducing size bias. We systematically compare four types of networks that vary in degree heterogeneity and clustering. We note that the RW and MHRW algorithms are based on Markov chains and have the "memoryless" property, leading to duplicate samples. Thus, we also compare and evaluate the effect of duplicate sample removal on the effectiveness of RW and MHRW. We assume undirected, static contact networks (Newman, 2018), implying that disease transmission can occur in both directions along a connection between two individuals and that the network structure remains unchanged over time.

We begin with Erdös-Rényi (ER) networks (Erdös and Rényi, 1959), which serve as a baseline for degree homogeneity and clustering. We then progress to Small-world (SW) networks (Watts and Strogatz, 1998), which introduce clustering and closely mimic social mixing networks. Finally, we consider the Negative-Binomial (NB) and Scale-free (SF) networks, which exhibit increasing levels of degree heterogeneity and are characterised by the presence of hubs with high connectivity, reflecting the presence of super-spreaders (Lieberthal and Gardner, 2021). The sequence of four network types, ER, SW, NB, and SF, is illustrated in Fig. 1, showing a progression from homogeneous to heterogeneous network structures. We also assume a stochastic Susceptible-Infected-Recovered (SIR) model (Brauer, 2008; Kermack and McKendrick, 1927) to simulate disease spread. We chose the SIR model because it captures heterogeneous mixing patterns within the population, as determined by the contact network structure (Kiss et al., 2017a).

Previous studies have shown that the RW algorithm can yield size-biased samples when applied to heterogeneous networks. In Gjoka et al. (2010), RW produces a size-biased sample of the Facebook network, over-representing high-degree nodes. In contrast, MHRW generates a sample with a degree distribution closely matching the UN. Similar observations are reported in Leskovec and Faloutsos (2006) for various network types, including citation networks from the e-print arXiv repository (Leskovec and Faloutsos, 2006), autonomous system networks (University of Oregon, 2025) (Internet router networks), affiliation networks in the Astrophysics category on arXiv (Leskovec and Faloutsos, 2006), and trust networks from Epinions.com (Richardson et al., 2003). However, a gap remains in the epidemiological literature, despite the widespread use of network sampling in disease modelling; a systematic comparison of RW and MHRW has not been performed. We set out to do this here. One of our aims is to provide clear recommendations for policymaking.

Although this is a technical paper, the results have high practical relevance. Policymakers and other readers interested in the main findings may refer directly to Section 3.2. For readers seeking technical details, we outline our methodology in Section 2 and present results in Section 3, highlighting the key trends. Additional results that reinforce these trends are provided in Appendices D and E.

The paper is structured as follows. In Section 2, we quantify size bias in the estimation of disease metrics for RW and MHRW, and in Section 3.2 we compare three disease metrics, the number of infected individuals, the number of secondary infections, and the time-to-infection, for the four chosen network types (ER, SW, NB, SF). To assess the robustness of MHRW in reducing size bias, in Section 3.3 we perform a sensitivity analysis by varying network and disease parameters. Finally, to illustrate the practical applications of our study, in Section 4.1 we analyse cattle movement network data provided by the British Cattle Movement Service (BCMS) (Personal Communication, June 14, 2024), which has hybrid properties related to the ER, SW and SF networks. We also analyse two human contact networks; in Section 4.2 data from residents of Haslemere town in England, UK (BBC, 2018), which is similar to the SW networks, and in Section 4.3 data from visitors of Science Gallery in Dublin, Ireland (Isella et al., 2011a), which is similar to the SW and NB networks. We provide a summary and directions for future research in Section 5.

2. Methods

In this section, we briefly describe the three main components of our methodology: (1) network structures, (2) sampling algorithms, and (3) a disease transmission model.

First, we generate four types of network: ER, SW, NB and SF (see Appendix A.1). We generate $n = 10,000$ realisations for each network type using the parameters listed in Table 1 to achieve a precision of 10^{-2} in the mean estimates of network characteristics. The value of n is based on the relationship between sample size (or the number of simulations) and the standard error of the mean (Bondy and Zlot, 1976), given by

$$\bar{\sigma}_x = \frac{\sigma}{\sqrt{n}}, \tag{1}$$

where $\bar{\sigma}_x$ is standard error of the mean, σ is the population standard deviation, and n is the sample size.

Second, we employ RW and MHRW sampling algorithms (see Appendix A.2) to sample from the networks. We obtain samples with 500 nodes for each network type (ER, SW, NB, and SF), using RW and MHRW (see pseudocode in Appendices 1,2). Here, the sample size of 500 (equivalent to 5 % of the population) is chosen to keep the margin of error approximately ±5 % in the disease spread estimates (Conroy et al., 2016).

Table 1
Chosen parameters for generating examples of four (ER, SW, NB, SF) networks using the NetworkX library (Hagberg et al., 2008).

Network parameter \ Network type	ER network	SW network	NB network	SF network
Number of nodes	10,000	10,000	10,000	10,000
Average degree	5	5	5	5
Edge creation Probability (p)	–	0.5	–	–
Dispersion parameter (r)	–	–	1	–
Power-law exponent (α)	–	–	–	3

However, we obtain 2500 nodes per sample and discard the initial 2000 nodes as burn-in nodes to allow the sampling algorithms to reach their stationary distribution and minimise the influence of the initial node on the resulting samples. Further, to get the precision of mean disease spread estimates up to the third decimal place (Bondy and Zlot, 1976) from the samples, we generate 100 samples (each with 500 nodes) from each of the 10,000 network realisations of four network types.

Finally, we model disease transmission using a stochastic SIR model (see Appendix A.3), simulating disease spread on each network with a Gillespie algorithm (Gillespie, 2007). The stochastic SIR model splits the population into three compartments: Susceptible, Infected, and Recovered. We assign rates with which individuals change compartments; β/N is the transmission rate for someone to become infected, and γ is the recovery rate (reciprocal of recovery period in days). We consider a range of values for β/N from 0 to 1, at an interval of 0.1. With $\gamma = 1$, β/N can be interpreted as a basic reproduction number. To understand the impact of longer recovery periods, we use $\gamma = 1/7$ and $1/14$, corresponding to average recovery periods for influenza (one week) (Chowell et al., 2008) and COVID-19 (two weeks) (SeyedAlinaghi et al., 2021), respectively.

For further implementation details, please refer to the code for the entire process pipeline on GitHub.

2.1. Hypothesis test

We perform hypothesis testing to statistically compare the estimates of disease metrics in RW and MHRW samples relative to the underlying network (UN). The first step is to check if the data follows a Normal distribution. We use a one-sample Kolmogorov-Smirnov (KS) test (Massey, 1951) to test the normality hypothesis for each of the four networks (ER, SW, NB and SF). We then use the one-tailed Mann-Whitney U (U) test (Mann and Whitney, 1947) to compare two samples and understand whether the samples are overestimating or underestimating disease spread metrics relative to the UN. We perform the hypothesis test for the ER, SW, NB, and SF networks, respectively (see Appendix A.5 for hypothesis statements).

3. Results

In this section, we compare the RW and MHRW sampling algorithms, assessing their effectiveness in sampling to capture the degree distribution of the UN and estimate the three disease metrics for four network types (ER, SW, NB, and SF).

Since RW and MHRW are memoryless algorithms, they yield duplicate nodes in a sample (Hu and Lau, 2013). We also compare the results when duplicates are removed or retained. This additional constraint aims to highlight how such a subtle aspect could significantly influence the disease metric estimates.

3.1. Degree distribution

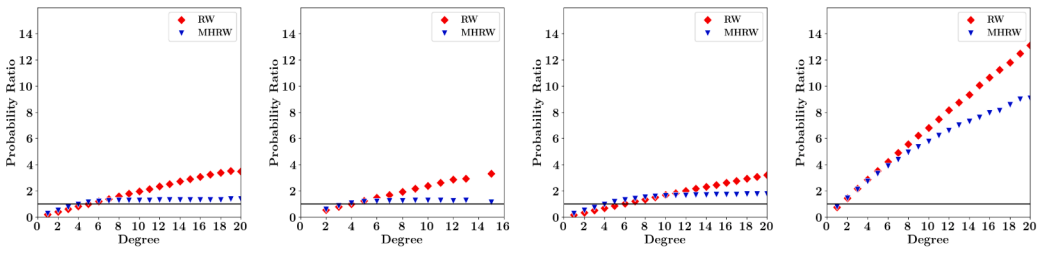
In Fig. 2, we present the ratio of the degree distributions from RW and MHRW samples compared to the UN, across ER, SW, NB, and SF networks. The vertical axis represents the ratio of the probability that a node has degree k in the sample relative to the UN. The solid black line represents a ratio of 1, so proximity to this line indicates that the sample degree distribution is representative of the UN degree distribution. Results are consistent whether duplicate nodes are included or excluded (data not shown).

For the ER (Fig. 2(a)) network, when the degree is ≥ 5, the MHRW samples (blue triangles) align closely with the UN. In contrast, the RW samples (red diamonds) increasingly deviate as the degree values increase, indicating size bias. This behaviour persists with the increase in clustering, i.e., for SW (Fig. 2(b)) networks. On the other hand, with an increase in degree heterogeneity from ER to NB networks (Fig. 2(c)), we observe a similar behaviour, but for degree values ≥ 10. Further increasing the heterogeneity of the network, SF networks (Fig. 2(d)), we observe that both sampling algorithms increasingly deviate from the UN with an increase in degree value. The increase in the dominance of "hubs", i.e., the high-degree nodes, as we move from NB to SF networks, leads to a high variance in the degree distribution and limits the MHRW's ability to reduce the size bias.

Our findings highlight that MHRW reduces size bias and provides a sample with a degree distribution similar to that of the UN, for networks with moderate heterogeneity and clustering. The networks with high variance in degree distribution, such as SF networks, require a more dedicated approach, such as Sampling for Large-Scale Networks at Low Sampling Rates (SLSR) (Jiao, 2024), where core and peripheral nodes are sampled separately for balanced sampling.

3.2. Disease metrics

In the previous section, we compared the degree distribution of the samples generated using the RW and MHRW algorithms. We observed that the degree distribution of MHRW samples is closer to that of the UN than that of RW samples due to a bias towards high-degree nodes.



(a) Erdős-Rényi network (b) Small-world network (c) Neg-binomial network (d) Scale-free network

Fig. 2. Samples obtained using the RW algorithm show size bias (over-representing high-degree nodes), MHRW algorithm reduces the size bias, resulting in a degree distribution closer to that of the UN. The ratio of degree distributions for RW (red diamonds) and MHRW (blue triangles) samples, relative to the UN, is presented for four network types (ER, SW, NB, SF). The UN degree distribution is based on 10,000 network realisations of each type (parameters are detailed in Table 1), generated using the NetworkX library (Hagberg et al., 2008). The degree distribution of samples from both sampling algorithms is estimated using 100 samples of size 500 for the 10,000 network realisations.

Now, we compare the sampling algorithms in estimating three disease spread metrics: 1) proportion of infected nodes, 2) average number of secondary infections, and 3) time-to-infection. We employ a stochastic SIR model to simulate disease transmission on networks.

Here, we present the results for $\gamma = 1$; see Appendix E.1 for $\gamma \in \{1/7, 1/14\}$ for ER and SW networks. We observe that as γ decreases, for both sampling algorithms, the size bias in estimates of the proportion of infected individuals decreases for ER and SW networks. Conversely, a decrease in γ increases the size bias in the number of secondary infections and time-to-infection estimates. Nonetheless, regardless of the γ value, estimates from the MHRW samples remain closer to the UN than those from the RW.

We also employ hypothesis testing to statistically compare disease spread estimates from the RW and MHRW sampling algorithms for the ER, SW, NB, and SF networks. As a first step, we apply a one-sample KS test to assess whether the distributions of three disease metrics follow a Normal distribution. The significance level is set at $\alpha = 0.05$. Results of the KS test are reported in Tables C.1–C.3. For each of the three disease metrics across four network types, for the UN, RW, and MHRW data, the p-values are consistently below 0.05. Thus, we reject the null hypothesis, i.e., data follows a Normal distribution, concluding that the data exhibit significant deviations from normality.

Since all three disease metrics deviate from a Normal distribution across the four network types, we employ a one-tailed U-test. The U-test compares the RW and MHRW estimates against the UN. The U-test statistic quantifies the similarity between the sample estimates and the UN, with larger U-statistic values indicating closer agreement. For each disease metric, we test two hypotheses: (i) statistically significant overestimation (Hypotheses 1, 3, 5), and (ii) statistically significant underestimation (Hypotheses 2, 4, 6). Thus, six hypotheses are evaluated over each of the three disease metrics, namely, the proportion of infected nodes, average number of secondary infections, and time-to-infection, across four network types. The significance level is set at $\alpha = 0.05$ for all hypothesis tests. The U-test results, along with comparative plots of RW and MHRW estimates relative to the UN, are presented in the following sections.

3.2.1. Proportion of infected nodes

In Fig. 3, we compare the proportion of infected nodes in the RW and MHRW samples relative to the UN as the transmission rate β increases, with a recovery rate $\gamma = 1$, for four networks (ER, SW, NB, and SF). Panel 1 (Fig. 3(a)–(d)) shows the results for samples with duplicate nodes, while Panel 2 (Fig. 3(e)–(h)) shows results for samples without duplicates. The closer the ratio is to the solid black line representing a ratio of 1, the closer the sample estimates are to those of the UN.

For samples with duplicate nodes (Panel 1), the RW consistently overestimates the proportion of infected nodes for the ER network (Fig. 3(a)) once $\beta > 0.2$, whereas the MHRW produces estimates that remain closer to the UN. Increasing clustering, from ER to SW (Fig. 3(b)), reduces the overestimation bias of RW, while the ratio distribution remains stationary for MHRW. In contrast, an increase in variance of degree distribution, from ER to NB network (Fig. 3(c)), amplifies the overestimation by RW, while MHRW remains closer to the UN. In SF networks (Fig. 3(d)), with extremely high variance in degree distribution, both RW and MHRW overestimate the proportion of infected nodes with high variability, consistent with the oversampling behaviour in the degree distribution (Fig. 2(d)). Our findings illustrate a key limitation of the MHRW algorithm in reducing size bias for highly heterogeneous networks.

For samples without duplicate nodes (Panel 2), the relative behaviour of RW and MHRW remains the same. However, removing duplicates reduces MHRW's accuracy in ER and SW networks but improves its accuracy in the NB network. Overall, the comparisons show that MHRW is generally closer to the UN in ER, SW, and NB networks, but both algorithms fail to provide reliable estimates for SF networks.

Table 2
U-test statistics (Hypotheses 1 and 2) for comparing the estimates of the proportion of infected nodes in the samples generated using the RW and the MHRW algorithms with the UN, with and without duplicate nodes.

Network \ Sample	With duplicate nodes		Without duplicate nodes	
	RW	**MHRW**	**RW**	**MHRW**
ER	5.07×10^{11}	5.90×10^{11}	5.11×10^{11}	5.38×10^{11}
SW	5.8×10^{11}	5.9×10^{11}	5.82×10^{11}	5.86×10^{11}
NB	3.2×10^{11}	5.3×10^{11}	3.23×10^{11}	5.42×10^{11}
SF	6.1×10^{10}	2.8×10^{10}	5.72×10^{10}	7.83×10^{9}

Hypothesis testing results (Table 2): To complement the above analysis, we conducted one-tailed U-tests to compare the RW and MHRW estimates against the UN estimates statistically. RW significantly overestimates the proportion of infected nodes across all four network types, with p-values < 0.05 for Hypothesis 1, regardless of whether duplicate nodes are included. MHRW exhibits network-dependent behaviour: with duplicate nodes, it significantly underestimates the proportion of infected nodes for ER, SW, and NB networks (p-values < 0.05 for Hypothesis 2); without duplicate nodes, MHRW significantly overestimates for the same networks (p-values < 0.05 for Hypothesis 1). In SF networks, MHRW consistently overestimates in both cases.

Although neither RW nor MHRW produces estimates that are statistically similar to those of the UN, the U-test statistic values indicate a relative closeness. For ER, SW, and NB networks, the U-test statistics for MHRW are greater than those for RW, confirming that MHRW estimates lie closer to the UN. For instance, in the ER network with duplicate nodes, the U-test statistic is 5.90×10^{11} for MHRW compared to 5.07×10^{11} for RW, implying that MHRW estimates are closer to the UN. In contrast, RW outperforms MHRW for SF networks, with higher U-test statistics indicating smaller deviations from the UN. For example, in SF networks, with duplicates, RW achieves 6.1×10^{10} versus 2.8×10^{10} for MHRW. These findings indicate that the MHRW significantly reduces size bias in networks characterised by moderate heterogeneity and clustering, such as ER, SW, and NB networks. However, its effectiveness diminishes in highly heterogeneous SF networks, where MHRW and RW algorithms produce highly variable results.

Furthermore, RW tends to overestimate the proportion of infected nodes because it tends to favour high-degree nodes (illustrated in Fig. 2), which have a greater likelihood of being infected.

3.2.2. Average secondary infections

In Fig. 4, we compare the number of secondary infections for RW and MHRW samples relative to the UN as the transmission rate β increases and a recovery rate $\gamma = 1$ for ER, SW, NB, and SF networks. Panel 1 (Fig. 4(a)–(d)) presents results when duplicate nodes are retained, while Panel 2 (Fig. 4(e)–(h)) shows results when duplicate nodes are removed.

For samples with duplicate nodes, when $\beta > 0.3$, RW overestimates the number of secondary infections for the ER network (Fig. 4(a)), whereas MHRW estimates are closer to the UN. Increasing clustering, from ER to SW network (Fig. 4(b)), reduces the overestimation bias of RW, while the ratio distribution remains stationary for MHRW. In contrast, increasing the variance of degree distribution, from ER to NB network (Fig. 4(c)), amplifies the overestimation by RW, while the MHRW estimates remain closer to the UN. For the SF network (Fig. 4(d)) with extremely high degree variance, both RW and MHRW overestimate the number of secondary infections, exhibiting high variability, which is consistent with observations for the degree distribution and proportion of infected nodes. Our findings again show the limitation of MHRW in reducing size bias for highly heterogeneous networks.

For samples without duplicate nodes (Panel 2), the relative behaviour of RW and MHRW remains similar. However, removing duplicates reduces MHRW's accuracy in ER and SW networks but improves its accuracy in the NB network. Overall, the comparisons show that MHRW

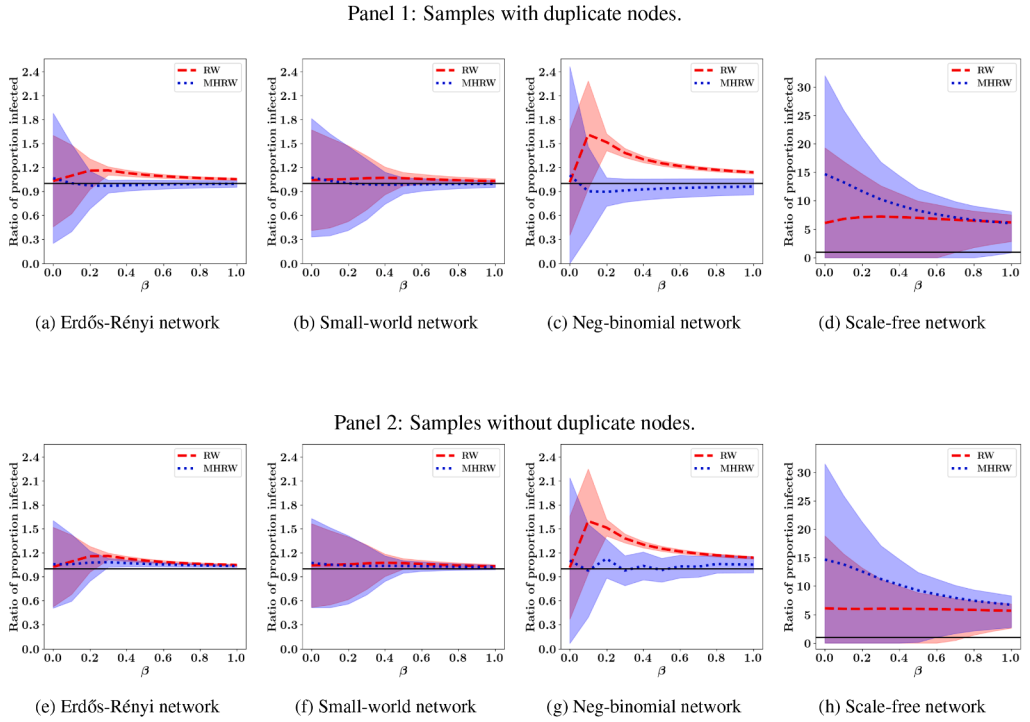


Fig. 3. The ratio of the MHRW estimates and the UN is closer to the black line, implying a reduction in size bias for ER, SW, and NB networks. The ratio of the proportion of infected nodes during an epidemic between RW (dashed red) and MHRW (dotted blue) samples relative to the UN for ER, SW, NB, and SF networks, is shown for retained (1) and removed duplicate nodes (2). The light red and blue bands indicate the range of one standard deviation. The UN infection count is based on SIR model simulations ($\gamma = 1$) for 10,000 networks of each type (the parameters are found in Table 1) generated with the NetworkX library (Hagberg et al., 2008). Sample infection counts are derived from 100 samples (size 500) for 10,000 network realisations using RW and MHRW.

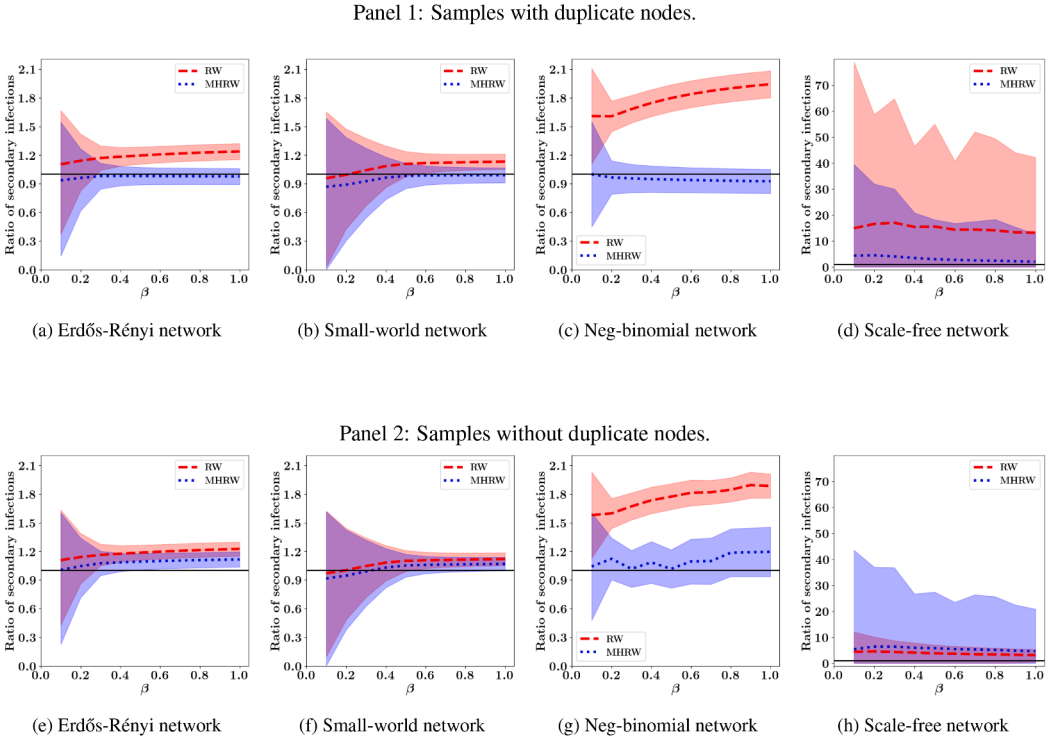


Fig. 4. The ratio of the MHRW estimates and UN is closer to the black line (1), implying better representation of the UN for ER, SW, and NB networks. The ratio of the number of secondary infections during an epidemic between RW (dashed red) and MHRW (dotted blue) samples, relative to the UN for ER, SW, NB, and SF networks, is shown for (1) retained and (2) removed duplicate nodes. The light red and blue bands indicate the range of one standard deviation. The UN estimates are based on SIR model simulations ($\gamma = 1$) for 10,000 network realisations of each type (the parameters are found in Table 1) generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples (size 500) for 10,000 network realisations using RW and MHRW.

Table 3
U-test statistics (Hypotheses 3 and 4) for comparing the estimates of the number of secondary infections in the samples generated using the RW and the MHRW algorithms with the UN, with and without duplicate nodes.

Network \ Sample	With duplicate nodes		Without duplicate nodes	
	RW	**MHRW**	**RW**	**MHRW**
ER	2.5×10^{11}	5.6×10^{11}	2.5×10^{11}	3.33×10^{11}
SW	3.99×10^{11}	6.0×10^{11}	3.91×10^{11}	4.51×10^{11}
NB	1.3×10^{11}	4.69×10^{11}	1.29×10^{11}	4.81×10^{11}
SF	8.2×10^{10}	6.6×10^{10}	8.45×10^{10}	5.99×10^{10}

Table 4
U-test statistics (Hypotheses 5 and 6) for comparing the estimates of the time-to-infection of the UN with the samples generated using the RW and the MHRW sampling algorithms, with and without duplicate nodes.

Network \ Sample	With duplicate nodes		Without duplicate nodes	
	RW	**MHRW**	**RW**	**MHRW**
ER	6.2×10^{11}	5.98×10^{11}	6.2×10^{11}	6.08×10^{11}
SW	6.1×10^{11}	6.0×10^{11}	3.91×10^{11}	4.51×10^{11}
NB	5.2×10^{11}	5.8×10^{11}	3.23×10^{11}	5.42×10^{11}
SF	1.26×10^{11}	7.37×10^{10}	8.45×10^{10}	5.99×10^{10}

is generally closer to the UN in ER, SW, and NB networks, but both algorithms fail to provide reliable estimates for SF networks.

Hypothesis testing results (Table 3):To complement the above analysis, we conducted one-tailed U-tests to compare the RW and MHRW estimates against the UN estimates statistically. RW samples, both with and without duplicates, significantly overestimate secondary infections across all four network types, with p-values < 0.05 providing evidence to reject H_0 in Hypothesis 3. MHRW samples without duplicates also significantly overestimate across networks (p-values < 0.05 for Hypothesis 3). With duplicates retained, MHRW significantly underestimates in ER, NB, and SF networks (rejecting H_0 for Hypothesis 4). But for the SW networks, the p-value > 0.05 for both Hypotheses 3, 4, indicating that MHRW estimates are statistically similar to the UN.

As with the proportion of infected nodes, the relative size of the U-test statistics indicates closeness to the UN. For ER, SW, and NB networks, MHRW consistently yields larger U-test statistics than RW (Table 3), confirming closer agreement with the UN. For example, for NB networks, without duplicate nodes, U-test statistics for MHRW is 4.81×10^{11} and for RW it is 1.29×10^{11}. However, in SF networks, RW estimates are closer to the UN, as higher U-test statistics indicate.

Our findings show that, similar to the proportion of infected nodes, MHRW significantly reduces size bias in networks with moderate heterogeneity and clustering, such as ER, SW, and NB networks. However, its effectiveness diminishes in highly heterogeneous SF networks, where MHRW and RW algorithms yield highly variable results. Additionally, the RW algorithm tends to overestimate the number of secondary infections due to bias towards high-degree nodes, which are more likely to be infected, as illustrated in Fig. 2.

3.2.3. Time-to-infection

In Fig. 5, we compare the ratio of the estimates of the time-to-infection from RW and MHRW samples relative to the UN as the transmission rate β increases and a recovery rate $\gamma = 1$ for ER, SW, NB, and SF networks. Panel 1 (Fig. 5(a)–(d)) shows results with duplicate nodes retained, while Panel 2 (Fig. 5(e)–(h)) shows the case with duplicate nodes removed.

For samples with duplicate nodes, when $\beta > 0.2$, the RW and MHRW estimates are closer to the UN across four network types and exhibit less bias compared to the previous two disease metrics: the proportion of infected nodes and the number of secondary infections. In ER networks (Fig. 5(a)), RW tend to underestimate the time-to-infection, while MHRW estimates are closer to the UN's. As clustering increases from ER to SW network (Fig. 5(b)), the underestimation bias of the RW decreases, whereas MHRW remains stable. Conversely, increasing the variance of the degree distribution from ER to NB network (Fig. 5(c)) amplifies the underestimation by RW, while MHRW continues to align closely with the UN. For the SF network (Fig. 5(d)), unlike the other two disease metrics, time-to-infection estimates for both sampling algorithms align well with the UN and show reduced variability.

For samples without duplicate nodes (Panel 2), the relative behaviour of RW and MHRW remains similar. However, removing duplicates reduces MHRW's accuracy for ER and SW networks but improves its accuracy for the NB network. Overall, the comparisons show that MHRW is generally closer to the UN in ER, SW, and NB networks, and both algorithms provide reliable estimates for SF networks.

Hypothesis testing results (Table 4): To complement the above analysis, we conducted one-tailed U-tests to compare the RW and MHRW estimates against the UN estimates statistically. RW samples (with and without duplicates) yield p-values < 0.05 for all four network types under Hypothesis 6, providing evidence of significant underestimation of the time-to-infection. For MHRW, estimates without duplicate nodes also consistently underestimate across all networks (p-values < 0.05 for Hypothesis 6), while with duplicate nodes retained, MHRW significantly underestimate in SW and SF networks (p-values < 0.05 for Hypothesis 6), but significantly overestimate in ER and NB networks (p-values < 0.05 for Hypothesis 5).

The U-test statistics in Table 4 indicate that in both ER and SW networks, RW and MHRW perform comparably regarding their closeness to the UN. However, in NB networks, the U-test statistics (both with and without duplicates) are higher for MHRW, suggesting that its estimates are closer to the UN compared to RW. Conversely, RW demonstrates higher U-test statistics in SF networks, indicating a closer alignment with the UN than MHRW.

Our findings indicate that, like the proportion of infected nodes and the number of secondary infections, MHRW effectively reduces size bias in moderately heterogeneous and clustered networks, such as ER, SW, and NB networks. Additionally, the RW algorithm tends to underestimate the time-to-infection due to its bias towards high-degree nodes (Fig. 2), which are more likely to be infected early in the epidemic.

3.3. Sensitivity analysis - network parameters

To examine the robustness of the sampling algorithms across four network types (ER, SW, NB, and SF), we perform a sensitivity analysis for a range of network parameters, listed in Table D.1. The sequence of networks considered reflects increasing structural complexity regarding clustering and degree heterogeneity. ER networks serve as a baseline of degree homogeneity and clustering, SW networks introduce clustering, and NB and SF networks introduce increasing levels of degree heterogeneity, as shown in Fig. 1. For each network type, we compare RW and MHRW algorithms for three disease spread metrics: 1) the proportion of infected nodes, 2) the number of secondary infections, and 3) the time-to-infection. Detailed results are provided in Appendix D. Below are the four key findings that arise from the sensitivity analyses:

1. The increase in network connectivity, achieved through either a larger network size or a higher average degree, consistently reduces bias for both sampling algorithms (RW and MHRW) across ER, SW, and NB networks.
2. Except for highly heterogeneous networks (SF networks), an increase in degree heterogeneity significantly penalises RW, resulting in higher size bias, while MHRW remains robust against this variation.
3. Except for highly clustered networks (Lattice networks), MHRW estimates consistently align more closely with the UN, whereas a small increase in bias is observed for RW estimates.

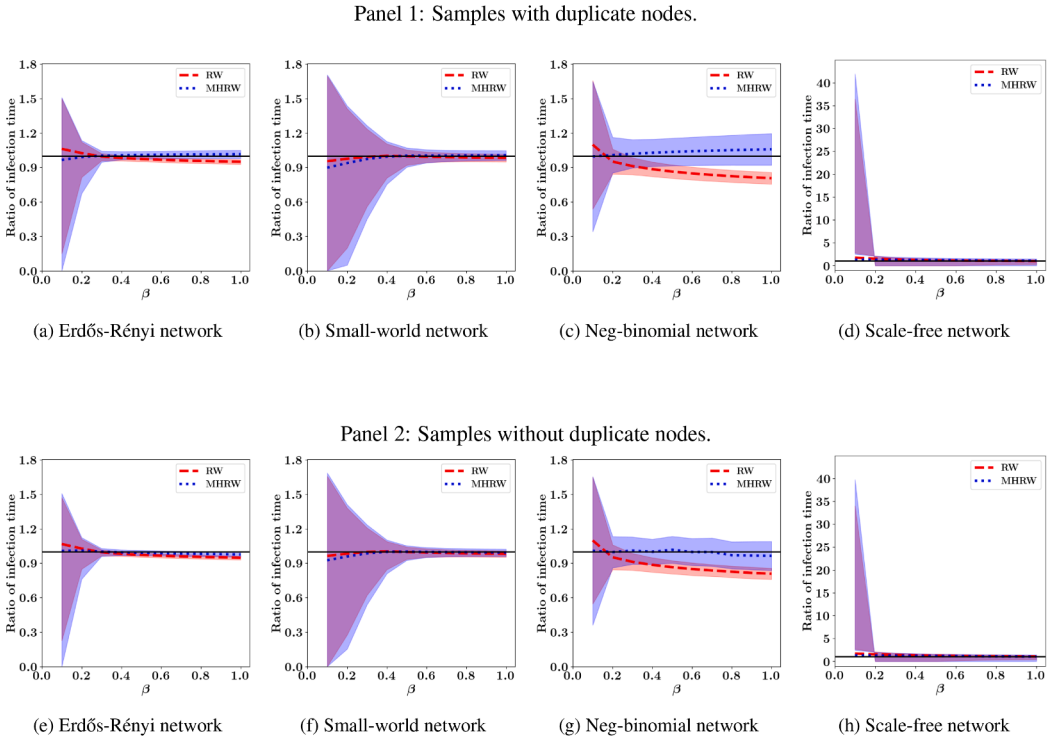


Fig. 5. The ratio of the MHRW estimates and UN is closer to the black line (1), implying better representation of the underlying network for ER, SW, NB, and SF networks. The ratio of the time-to-infection during an epidemic between RW (dashed red) and MHRW (dotted blue) samples, relative to their respective UNs, is shown for (1) retained and (2) removed duplicate nodes. The light red and blue bands indicate the range of one standard deviation. The UN estimates are based on SIR model simulations ($\gamma = 1$) for $10,000$ network realisations of each type (the parameters are found in Table 1) generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples (size 500) for $10,000$ network realisations using RW and MHRW.

4. The MHRW estimates are sensitive to the disease transmission parameter (β). Above a specific β threshold, MHRW estimates are closer to the UN, while below this threshold, MHRW estimates exhibit bias, and RW estimates remain biased regardless of the threshold. Additionally, this β threshold decreases with an increase in network size and average degree.

In summary, results demonstrate that while RW can perform adequately in homogeneous, well-connected networks such as ER and SW networks, MHRW offers more reliable performance across a broader spectrum of network structures, including ER, SW, and NB networks. Additionally, neither sampling algorithm works for highly clustered or highly heterogeneous networks.

4. Real-world networks

In Sections 3.2 and 3.3, we investigated the performance of the MHRW algorithm in reducing size bias in estimates of three disease spread metrics, compared to the RW algorithm, across four theoretical network types (ER, SW, NB, and SF). We also discussed the effect of variation in network parameters and disease transmission rate β on the bias in estimates from both the sampling algorithms.

In this section, we implement RW and MHRW algorithms on real-world networks. We study three networks: 1) cattle movement network from the British Cattle Movement Service (BCMS), 2) contact network data of residents of Haslemere town in England, UK, BBC (2018) and 3) face-to-face interaction data of the visitors at the Science Gallery in Dublin, Ireland (Isella et al., 2011a). All three network datasets are collected over a period; hence, the network is temporal, meaning that network connections and the number of nodes change over time. For this study, we create static networks from these datasets' snapshots to only study the impact of sampling algorithms.

4.1. Cattle movement network

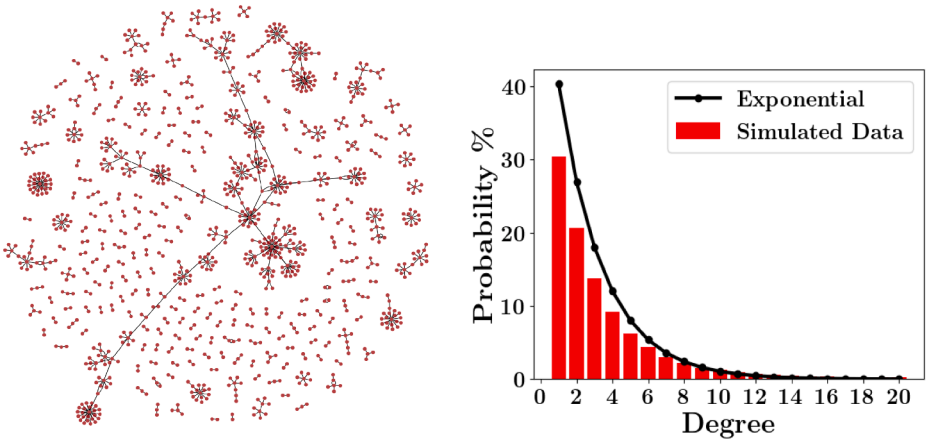
In this section, we implement RW and MHRW algorithms on cattle movement network data, provided by the British Cattle Movement Service (BCMS). This network data contains information on the birth, death, and movement of cattle. We use cattle movement data from January 2018 to March 2018, recorded at the day level. We consider cattle holding places (farms, marketplaces, and slaughterhouses) as nodes and an edge is created between two nodes if cattle movement exists between them. Note that movement does not occur through all edges daily, implying that the network structure changes over time. However, for this study, we assume that all edges are active and that the network remains static.

4.1.1. Data description

The data contains movement details of $1,135,502$ cattle; on average, there are two movement records per cattle. There are $46,512$ holding places, which are the nodes of the network. A static and undirected network is created with $46,512$ nodes and $159,036$ edges. Note that an edge between a pair of nodes is counted only once in the case of multiple movement records. A part of the movement network is shown in Fig. 6(a); we observe a mix of large connected components, small clusters of nodes, and isolated node pairs.

The network's degree ranges from 1 to 6556, with an average degree of 6. Notably, 30 % of the nodes have a degree of 1, while only 2.69 % of the nodes have a degree greater than 20. There is only one node with 6556 edges, which should be a marketplace.

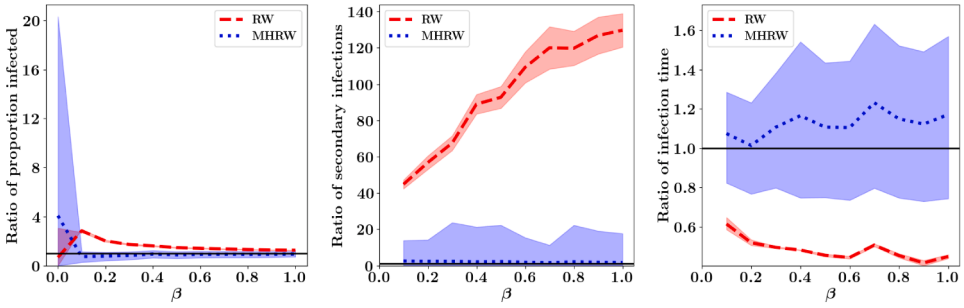
The degree distribution for the network is shown in Fig. 6(b), up to a degree value of 20. The degree distribution of the cattle network is fitted with an exponential distribution (Balakrishnan, 2019) with rate parameter 0.4 and location parameter 1 (the distribution is shifted by one unit to the right along the horizontal axis), represented by the solid black



(a) Subset of the cattle movement network. (b) Degree distribution.

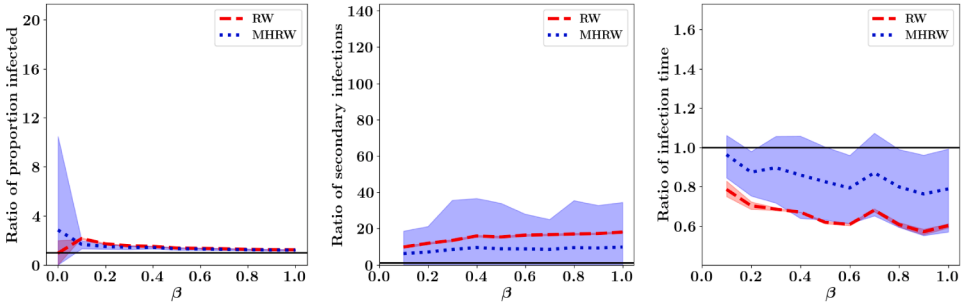
Fig. 6. An illustration of the cattle movement network data from BCMS exhibiting properties of ER and SW networks, and follows an exponential degree distribution. (a) A subset of the cattle movement network, where nodes represent individual farms and edges indicate recorded cattle movements between them. (b) Degree distribution of the cattle network. The red histogram represents the empirical data, and the black curve shows an exponential fit for comparison.

Panel 1: Samples with duplicate nodes.



(a) Proportion of infected nodes (b) Number of secondary infections (c) Time-to-infection (days)

Panel 2: Samples without duplicate nodes.



(d) Proportion of infected nodes (e) Number of secondary infections (f) Time-to-infection (days)

Fig. 7. The ratio of the MHRW sample estimates and UN is closer to the black line for all three disease metrics compared to RW, implying better representation of the cattle movement network. The ratio of disease metric estimates between samples generated using the RW and MHRW algorithms, relative to the UN, for two scenarios: (1) samples with duplicate nodes, and (2) samples without duplicate nodes. The light red and blue bands indicate the range of one standard deviation. Disease metric estimates are based on 500 SIR model simulations on the cattle movement network with $\gamma = 1$. Sample estimates are derived from 10,000 samples (size 2500) for the cattle movement network, using RW and MHRW algorithms.

line in Fig. 6(b). Other studies state that the cattle movement network follows a power-law degree distribution (Christley et al., 2005; Fielding et al., 2019). One reason for this difference is the amount and period of data; we have only three months of data (from 2018), whereas these studies analyse data for multiple years preceding 2018 (Duncan et al., 2022).

We observe clusters with few connections outside them, similar to the structure of the SW networks and consistent with the results in Fielding et al. (2019). We also observe a uniform connectivity pattern (average degree 2), similar to the SW and ER networks, if we exclude nodes with one connection or a degree greater than 1000. We observe NB and SF networks' trait of the presence of "hubs" in the cattle movement

network, which is suggested by the presence of 0.06 % nodes with degree ≥ 1000 and 30 % nodes with degree 1. There are a few nodes with very high degrees, likely markets, that facilitate extensive animal movement and can potentially act as key nodes for disease spread.

4.1.2. Results

Due to a lack of disease data, we study a hypothetical disease spread scenario, for example, bovine tuberculosis (Bekara et al., 2014). We run 500 simulations of the stochastic SIR model on the cattle movement network using the methodology and parameters from Section 2. We use the RW and MHRW sampling algorithms to generate 10,000 samples, with a sample size of 2500 (5 % of the network size). The number of samples, sample size, and number of SIR model simulations are chosen to be accurate to the second decimal place, as in Section 3.

In Fig. 7, we compare estimates of three disease metrics from RW and MHRW samples relative to the cattle network, as the transmission rate β increases, with a recovery rate $\gamma = 1$. The three disease metrics are: 1) proportion of infected nodes, 2) number of secondary infections, and 3) time-to-infection. Panel 1 (Fig. 7(a)–(c)) shows the results for samples with duplicate nodes, while Panel 2 (Fig. 7(d)–(f)) shows those without duplicate nodes. The closer the disease metric estimates are to the solid horizontal black line, the better the agreement with the metric values of the underlying cattle network.

We observe that RW overestimates the proportion of infected nodes in both panels (with and without duplicate nodes); see Fig. 7a and d. However, the estimate from MHRW samples aligns very well with the cattle network when duplicate nodes are retained (Fig. 7(a)), but it overestimates when duplicate nodes are removed (Fig. 7(d)). Additionally, we find that the RW and MHRW estimates overlap when $\beta > 0.3$ for samples without duplicate nodes, which aligns with the observations in ER and SW networks (Fig. 3).

For the number of secondary infections (Fig. 7(b), (e)), we observe significant overestimation by RW for both cases with (by a factor of 40-130) and without duplicate nodes (by a factor of 10-20). In comparison, MHRW estimates remain closer to the cattle network values. Specifically, when duplicate nodes are retained, MHRW estimates align perfectly with the cattle network, but when duplicate nodes are removed, we observe an overestimation by a factor of ~ 9. Again, these observations align with those of ER, SW, and NB networks (Fig. 4).

Both sampling algorithms fail to estimate the time-to-infection (Fig. 7(c), (f). RW underestimates time-to-infection, and the bias decreases when duplicate nodes are removed. On the other hand, MHRW overestimates time-to-infection when duplicates are retained and underestimates it when duplicates are removed, but has less bias compared to the RW estimates. This pattern of RW and MHRW sample estimates is similar to that of NB networks, as shown in Fig. 5.

For the cattle movement network, although MHRW estimates are closer to the underlying network due to a reduction in size bias, the variation in estimates is higher for MHRW compared to RW, specifically for the number of secondary infections and time-to-infection. We do not find this in the ER, SW, and NB network results. The high variance in estimates is potentially due to the presence of segregated clusters and the isolated pair of nodes, leading to challenges in traversing the majority of the cattle network.

The trend of disease metric estimates on the cattle movement network is similar to that in the ER, SW, and NB networks. RW provides conservative estimates of disease spread, whereas the MHRW yields estimates closer to the UN when duplicate nodes are retained. If the cattle movement network is similar to the one in this study, for highly contagious diseases, such as Foot and Mouth Disease, which has a mortality rate of 20 % in young calves (World Organisation for Animal Health, 2024) and an average of 20 secondary infections for direct transmission (Paton et al., 2018), the RW sampling algorithm is recommended. Conversely, for diseases such as Bovine Viral Diarrhoea, with a mortality rate of approximately 7 % (Dobos et al., 2024) and the average number of secondary infections as 0.36 (Isoda et al., 2025), the MHRW algorithm is preferable as it provides more representative estimates, which is particularly valuable for optimising intervention policies.

4.2. Haslemere residents' contact network

In this section, we implement RW and MHRW algorithms on the Haslemere contact network dataset collected as part of the BBC documentary "Contagion! The BBC Four Pandemic" (BBC, 2018). It captures social interaction data from residents of Haslemere, a town in England, UK. It comprises proximity records for 468 individuals over three consecutive days, with location tracked at five-minute intervals with an approximate resolution of one meter. A social contact is defined as an event when two individuals are within four meters of each other for a 5 min period; this threshold is comparable to Bluetooth-based proximity tracing and is chosen to reflect face-to-face contact relevant for transmission modelling. This high-resolution dataset includes 1257 unique social links and 1616 daily contact events, offering a representative view of daily contact patterns in a local setting. Please see Firth et al. (2020) for a detailed description of the contact network.

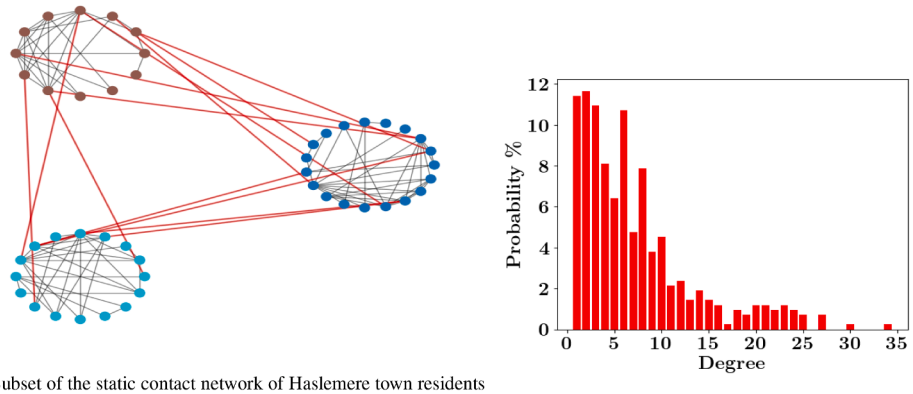
We construct a static network from the three-day dataset using data from 13th October 2017, the day with the highest number of participants and the largest connected component. The network of 13th October 2017 consists of 420 individuals represented as nodes and 1506 social links as edges. We assume that all social connections recorded during the day are active and that the network remains static throughout this period. Additionally, all edges are considered to have equal weight and are active for the entire day. We observe the presence of long-range connections and tightly knit groups, such as households, which are characteristics of SW networks (Fig. 8).

Due to a lack of disease data, we study a hypothetical disease spread scenario. We run 500 simulations of the stochastic SIR model on the Haslemere network using the methodology and parameters from Section 2. We generate 5000 samples, with a sample size of 21 (5 % of the network size) using the RW and MHRW algorithms. The number of samples, sample size, and number of SIR model simulations are chosen to be accurate to the second decimal place, as described in Section 2.

In Fig. 9, we compare estimates of three disease metrics: the proportion of infected nodes, number of secondary infections, and time-to-infection, from RW and MHRW samples relative to the estimates from the underlying Haslemere network. We consider a fixed recovery rate $\gamma = 1$ and vary the transmission rate β. Panel 1 (Fig. 9(a)–(c)) shows the results for samples with duplicate nodes, while Panel 2 (Fig. 9(d)–(f)) shows those without duplicates. The closer the disease metric estimates are to the solid horizontal black line, the better the agreement with the Haslemere network.

We observe that the estimates of the proportion of infected nodes by the RW and MHRW algorithms align with the Haslemere network for $\beta \geq 0.2$, in both cases with and without duplicate nodes. Note the β threshold pattern as discussed in Section 3.3. We find that RW overestimates the number of secondary infections by more than 50 % in both cases, whether or not duplicate nodes are present. In contrast, MHRW estimates are closer to the Haslemere network, with approximately a 10 % underestimation when duplicate nodes are retained and a 10 % overestimation when duplicate nodes are removed. The time-to-infection estimates align with the Haslemere networks for MHRW when duplicates are removed, whereas MHRW overestimates by 5 % when duplicates are retained. RW underestimates the time-to-infection by 10 % for both cases, with or without duplicates.

The analysis of the Haslemere network shows that the trends and patterns in estimates of the number of secondary infections and the time-to-infection closely align with those observed in the NB networks. Additionally, the proportion of infected nodes follows a pattern similar to that of SW networks. In the context of less fatal diseases such as seasonal influenza ($\beta = 0.3$) (Chowell et al., 2008), it is recommended to utilise the MHRW algorithm to obtain precise estimations, particularly if the contact network is similar to that of the Haslemere network.

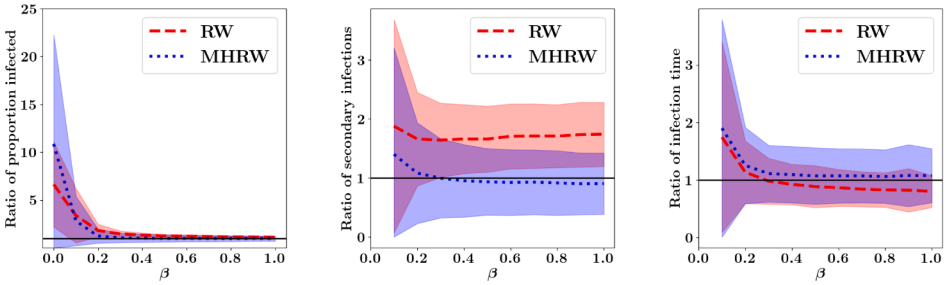


(a) Subset of the static contact network of Haslemere town residents on 13th October 2017.

(b) Degree distribution

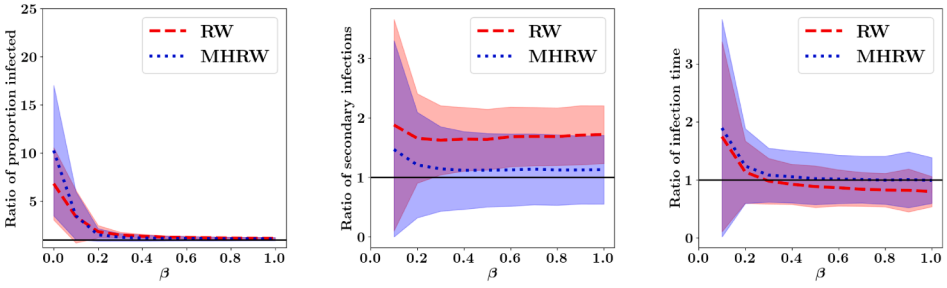
Fig. 8. Illustration of a subset of the contact network data collected on 13th October 2017 in Haslemere town (BBC, 2018). (a) A subset of the contact network of Haslemere residents, where nodes represent town residents and edges indicate an interaction between two residents. Nodes of the same colour belong to a community, where grey colour edges represent within-community connections and red colour edges represent the long-range connections to other communities. (b) Degree distribution of the contact network of Haslemere town.

Panel 1: Samples with duplicate nodes.



(a) Proportion of infected nodes (b) Number of secondary infections (c) Time-to-infection (days)

Panel 2: Samples without duplicate nodes.



(d) Proportion of infected nodes (e) Number of secondary infections (f) Time-to-infection (days)

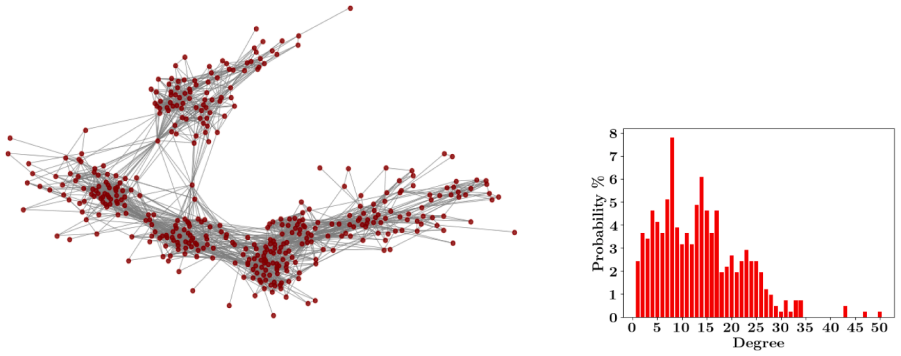
Fig. 9. The ratio of MHRW estimates and UN is closer to the black line for all three disease metrics, implying better representation of the Haslemere network. The ratio of disease metric estimates between samples generated using the RW and MHRW algorithms, relative to the UN, for two scenarios: (1) samples with duplicate nodes, and (2) samples without duplicate nodes. The light red and blue bands indicate the range of one standard deviation. Disease metric estimates are based on 500 SIR model simulations on the Haslemere network with $\gamma = 1$. Sample estimates are derived from 5000 samples (size 21) for the Haslemere network, using the RW and MHRW algorithms.

Conversely, for highly contagious and fatal diseases like Measles ($\beta = 0.6$) (Guerra et al., 2017), it is crucial to use the RW algorithm to minimise the risk of fatalities and reduce mortality rates.

4.3. Science gallery visitors' interaction network

The Infectious SocioPatterns dynamic contact networks dataset captures detailed temporal face-to-face interactions collected during the "INFECTIOUS: STAY AWAY" art science exhibition at the Science Gallery in Dublin, Ireland (Isella et al., 2011a). The data record contains information about the active contacts among anonymous participants, each lasting 20 s. The dataset provides daily dynamic contact networks representing very fine-grained temporal proximity patterns, suitable for studying dynamic social interactions and contagion processes. For a detailed day and time level data description, please see Isella et al. (2011a,b).

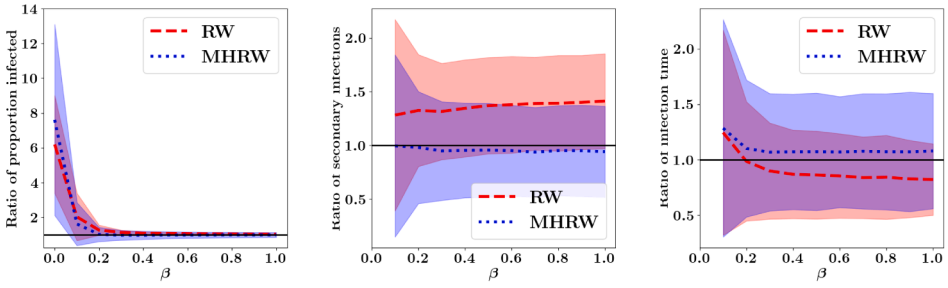
Similar to the Haslemere data, we consider the day with the maximum number of participants and the largest connected component to



(a) Science Gallery contact network on 15th July 2009. (b) Degree distribution

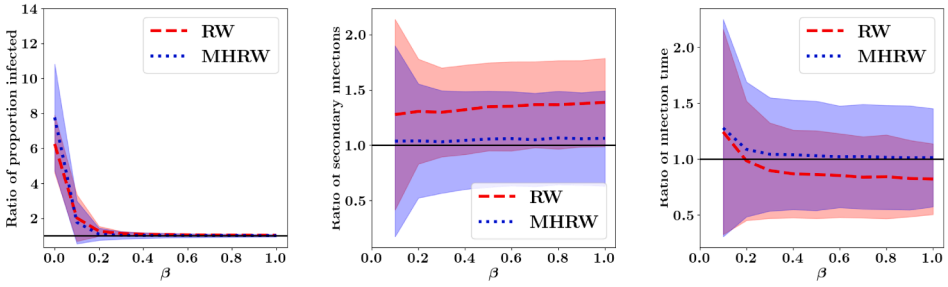
Fig. 10. Illustration of the contact network data of visitors at Science Gallery on 15th July 2009. (a) Face-to-face interaction network, where nodes represent gallery visitors and edges indicate an interaction between two visitors. (b) Degree distribution of the contact network at Science Gallery.

Panel 1: Samples with duplicate nodes.



(a) Proportion of infected nodes (b) Number of secondary infections (c) Time-to-infection (days)

Panel 2: Samples without duplicate nodes.



(d) Proportion of infected nodes (e) Number of secondary infections (f) Time-to-infection (days)

Fig. 11. The ratio of estimates from MHRW samples and the UN is closer to the black line for all three disease metrics, implying better representation of the Science Gallery network. The ratio of disease metric estimates from samples generated using RW and MHRW algorithms, relative to the UN, for two scenarios: (1) samples with duplicate nodes, and (2) samples without duplicate nodes. The light red and blue bands indicate the range of one standard deviation. Disease metric estimates are based on 500 SIR model simulations on the Science Gallery network with $\gamma = 1$. Sample estimates are derived from 5000 samples (size 21) for the Science Gallery network, using RW and MHRW.

study the worst-case scenario. We create a static network using the face-to-face interaction data collected at the Science Gallery on July 15, 2009, which consists of 410 individuals as nodes and 2765 social links as edges. We assume that all edges have the same weight and remain active at all times.

The Science Gallery networks show characteristics similar to those of SW and NB networks. For example, the presence of the cluster of nodes in the network (Fig. 10) with some connections in between the clusters, like SW networks. The presence of nodes with a high number of connections (> 40), as shown in Fig. 10, is a characteristic of NB networks. Unlike Haslemere (Fisher-Pearson skewness coefficient (Doane and Seward, 2011) of 1.53) and Cattle movement networks (Fisher-Pearson skewness coefficient of 49.71), we find that the degree distribution (Fig. 10(b)) is less skewed to the right, with a Fisher-Pearson skewness coefficient of 0.86, making the network relatively less heterogeneous.

Due to a lack of disease data, we study a hypothetical disease spread scenario. We run 500 simulations of the stochastic SIR model on the Science Gallery network using the methodology and parameters from Section 2. We use RW and MHRW algorithms to generate 5000 samples,

with a sample size of 21 (5 % of the network size). The number of samples, sample size, and number of SIR model simulations are chosen to be accurate to the second decimal place, as in Section 2.

In Fig. 11, we compare estimates of three disease metrics: the proportion of infected nodes, number of secondary infections, and time-to-infection, from RW and MHRW samples relative to the estimates from the underlying Science Gallery network. We consider a fixed recovery rate $\gamma = 1$ and vary the transmission rate β. Panel 1 (Fig. 11(a)–(c)) shows the results for samples with duplicate nodes, while Panel 2 (Fig. 11(d)–(f)) shows those without duplicates. The closer the disease metric estimates are to the solid horizontal black line, the better the agreement with the underlying Science Gallery network.

For the proportion of infected nodes, the estimates from both sampling algorithms are close to each other and align with the underlying Science Gallery network estimates for $\beta > 0.2$, in both panels (with and without duplicate nodes). For $\beta \leq 0.2$, both sampling algorithms significantly overestimate the proportion of infected nodes.

The RW overestimates the number of secondary infections by 40 % in both cases with and without duplicate nodes, when $\beta > 0.2$. In contrast, MHRW estimates are closer to the underlying Science Gallery network, with 10 % underestimation when duplicates are retained and 10 % overestimation when duplicates are removed. These findings are the same as those of the Haslemere network (Section 4.2) and the NB networks (Section 3.2.2).

For time-to-infection, the RW underestimates by 10 % for $\beta > 0.2$ in both panels (with and without duplicate nodes). In comparison, the MHRW overestimates by 5 % for $\beta > 0.2$ when duplicates are retained and aligns well with the Science Gallery network when duplicates are removed. Again, we observe a similarity with the results of the Haslemere network (Section 4.2).

These findings demonstrate that for the Science Gallery network, MHRW can reduce size bias in the number of secondary infections and time-to-infection and provides estimates comparable to RW for the proportion of infected nodes. Overall, if the network is similar to the Science Gallery and the disease is severe, such as Smallpox, $\beta = 0.9$ (Gani and Leach, 2001), it is advised to use RW sampling to get conservative estimates and reduce the mortality rate. Whereas for less severe diseases, such as Seasonal influenza, $\beta = 0.3$ (Guerra et al., 2017), MHRW is a better sampling algorithm for optimising resource allocation with precise disease spread metric estimations.

5. Conclusions and further directions

In this study, we evaluated the performance of the Metropolis-Hastings Random Walk (MHRW) relative to the Random Walk (RW) sampling algorithm in reducing size bias, the over-representation of highly connected individuals, when estimating disease spread metrics across four network types: Erdös-Rényi (ER), Small-world (SW), Negative-binomial (NB), and Scale-free (SF). These networks were selected to reflect increasing structural complexity, from the degree homogeneity and low clustering of ER networks to the clustering introduced by SW networks, and the degree heterogeneity of NB and SF networks. Together, they capture essential features of real-world contact networks.

For example, sexual contact networks have been shown to resemble ER networks (Holme, 2013), COVID-19 transmission network in Houston, Texas aligns more closely with SF networks (Fujimoto et al., 2023), rabies spread among Serengeti lions has parallels with SW networks (Craft et al., 2011), and human social networks are often best represented by NB networks with strong community structures (Newman, 2006).

We employed a stochastic Susceptible-Infected-Recovered (SIR) model to simulate disease transmission on networks. We assessed whether the MHRW algorithm can reduce size bias that skews disease metric estimates when using the RW sampling algorithm. Our evaluation centred on three key disease spread metrics: proportion of infected nodes, number of secondary infections, and time-to-infection. We further examined the role of duplicate nodes on bias in the estimates of the disease spread metrics.

Our analyses (Sections 3.2, 3.3) demonstrate that the MHRW algorithm provides more precise estimates of three disease metrics than the RW algorithm for networks with moderate clustering and heterogeneity (ER, SW, and NB). Increasing network size and connectivity (average degree) consistently reduced size bias for both sampling algorithms. However, RW estimates are strongly penalised by increased degree heterogeneity, whereas MHRW estimates remain robust. On the other hand, an increase in clustering led to a small increase in size bias for RW samples, whereas MHRW estimates remain stable and robust again. MHRW estimates aligned more closely with the underlying network except in two extreme cases: highly clustered lattice networks and highly heterogeneous SF networks, where both algorithms failed to provide representative samples.

Our findings suggest that RW is suitable for rapid, early assessments in severe, fast-spreading epidemics for networks similar to the ER and SW. In comparison, MHRW is better suited for slower, lower-severity epidemics where precision in estimates is more critical, for moderately clustered and heterogeneous networks (such as ER, SW, and NB networks). For networks with extreme clustering or heterogeneity, alternative sampling strategies may be necessary, such as hybrid approaches or algorithms like SLSR (Jiao, 2024). Overall, MHRW expands the range of network structures where size bias can be effectively reduced, enhancing the accuracy of epidemic modelling on complex networks.

Our analysis also revealed that disease parameters significantly impact the bias in estimates of the disease spread metrics. As recovery rate (γ) increases, the difference between RW and MHRW estimates of the proportion of infected nodes (Appendix E.1) decreases. Moreover, above a transmission rate (β) threshold, MHRW estimates align almost perfectly with the underlying network (Appendix D.4), whereas RW estimates remain biased.

In our analysis of real-world networks (Section 4), we observed patterns consistent with the theoretical network results. For the cattle movement network, which combines ER, SW, and NB network features, MHRW produced estimates that more accurately reflected the underlying network, particularly when duplicate nodes were retained. By contrast, RW sampling introduced a substantial size bias, leading to an overestimation of the proportion of infected nodes, a significant overestimation of the number of secondary infections, and an underestimation of the time-to-infection. These results highlight RW's tendency toward conservative estimates, valuable for rapid responses in highly contagious outbreaks, such as foot-and-mouth disease. In contrast, the more precise estimates from MHRW are better suited for resource-optimised interventions in slower-spreading epidemics, such as Bovine Viral Diarrhoea.

The analysis of real human contact networks further supports our findings. In the Haslemere residents' network, which resembles an SW network, MHRW consistently reduced size bias across all three disease spread metrics. In the Science Gallery visitors' network, which combines SW and NB network properties, MHRW reduced the size bias for all three disease spread metrics. Overall, the analysis of real-world networks confirms that while RW has limited utility for severe diseases on homogeneously connected networks such as ER and SW, MHRW offers more precise estimates for a wide range of contact networks, including ER, SW, and NB networks.

In this study, we demonstrated the effectiveness of the MHRW algorithm over the RW algorithm in reducing size bias when estimating disease metrics for four key network types. However, it is essential to note that MHRW, despite its advantages in reducing size bias, has limitations in networks with high-degree variance, such as SF networks as well as in highly clustered networks, like lattice networks. Additionally, MHRW relies on the assumption that node degrees are known or accessible during sampling, which may not always be feasible in dynamic networks. While retaining duplicate nodes improves the accuracy of the MHRW estimates, it increases computational costs and may reduce the scalability of the method in larger and more complex networks.

The use of the Gillespie algorithm for simulating stochastic SIR disease dynamics introduces certain constraints. It can become computationally expensive for larger networks, limiting the ability to perform large-scale simulations. Additionally, the choice of parameter values for transmission and recovery rates may not fully capture the diversity of epidemic dynamics in different settings. Although we selected parameters that reflect realistic disease scenarios, the variability in disease dynamics across various populations and network structures may affect the generalizability of our results.

Despite these limitations, our findings provide valuable insights into the behaviour of the RW and MHRW sampling algorithms and suggest areas for further work. To further understand the subtle differences between the RW and MHRW sampling algorithms with respect to network structure, it would be valuable to explore the impact of selecting a specific starting node for sampling rather than a random node, particularly for SF networks. Furthermore, as observed in Fig. E.1, the difference in disease estimates between the RW and MHRW samples decreases with the recovery rate. It would be interesting to determine the value of the recovery rate at which this change begins, as it could guide the choice of sampling algorithm.

CRediT authorship contribution statement

Neha Bansal: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Katerina Kaouri:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Thomas E. Woolley:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Data and code

Data and code are available here: https://github.com/nehabansal26/Epidemics-on-Networks

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work is supported by the Natural Environment, Biotechnology and Biological Sciences and Medical Research councils (NERC, BBSRC and MRC) [grant number: NE/X016714/1] as part of the One Health for One Environment: an A-Z Approach for Tackling Zoonoses ('OneZoo') Centre for Doctoral Training. This work is also supported by an EPSRC Impact Acceleration Account grant, under grant number EP/X525522/1.

Appendix A. Background

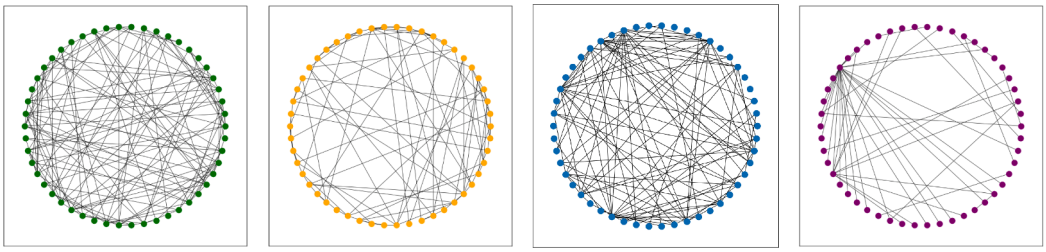
A.1. Networks

We consider contact networks (Bansal et al., 2010; Craft, 2015). A contact network is defined by $G = \{V, E\}$, where V is a set of node labels representing the individuals who can carry and transmit the disease, and E is a set of edges that define whether there have been interactions between the individuals. Let N be the number of nodes in G, also known as the network size. Let k_v denote the degree of a node $v \in V$, which is defined as the number of edges (connections) of node v. Based on the network structure, we can define and calculate (at least numerically) p_k, the probability that a node is of degree k, as well as the average degree of the network $\langle k \rangle$ (Newman, 2018). We will randomly generate our network structures based on these degree distributions. Namely, several nodes will be fixed, and the connections will be chosen to satisfy a given degree distribution.

A.1.1. Network structures

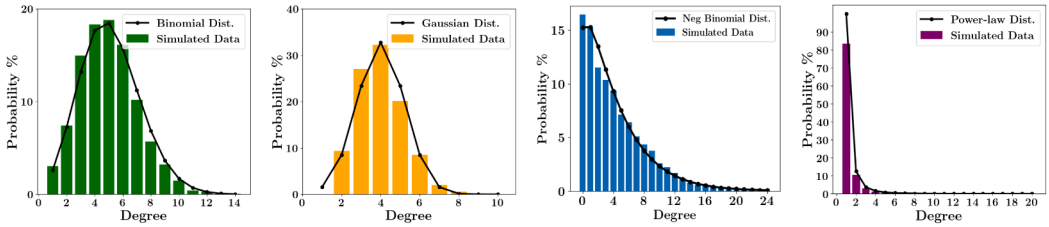
Real-world networks, such as the cattle movement network from the BCMS, and human contact networks of Haslemere town and the Science Gallery, which we analyse in Section 4, rarely align perfectly with one of the theoretical network structures like ER, SW, NB, or SF networks. Instead, they often have a mixture of features from these networks, such as clustering and heterogeneous connectivity. To better understand the implications of these structural characteristics on disease spread, we first describe these four theoretical networks individually (Newman, 2018). This approach helps us to determine how specific network features influence disease metrics, offering a baseline for understanding the more complex structure of the cattle movement network and human contact networks.

1. **Erdös-Rényi (ER) networks** are generated by considering every pair of nodes and connecting them with an edge with probability p, independently of other edges in the network (Erdös et al., 1960). Fig. A.1(a) shows an example of an ER network. ER networks follow binomial degree distribution (Newman et al., 2001), with network size N and edge creation probability $p = \langle k \rangle/(N-1)$ (Fig. A.2(a)), that is, given by
$$p_k = \binom{N}{k} p^k (1-p)^{N-k}. \tag{A.1}$$
2. **Small-world (SW) networks** are characterised by high clustering and a small average shortest path distance between nodes (see Fig. A.1(b)), and they follow a skewed Gaussian degree distribution (Fig. A.2(b)). Two common methods for generating SW networks are: 1) the Watts-Strogatz (WS) model (Watts and Strogatz, 1998), and 2) the Newman-Watts-Strogatz (NWS) model (Newman and Watts, 1999). Both methods begin with a regular lattice network with average degree $\langle k \rangle$. In the NWS method, edges are added with probability p; in the WS method, edges are rewired with probability p. For $p = 0$, the network is a regular lattice, and for $p = 1$, the network is an ER network. For $0 < p < 1$, a range of small-world networks with varying characteristics can be generated, and the value of p can be selected based on the case. The example of an SW network shown in Fig. A.1(b) is generated using the WS method.
3. **Negative-binomial (NB) networks** (see Fig. A.1(c)) follow a negative-binomial degree distribution, characterised by a heavy tail in the distribution, i.e., the presence of nodes with a relatively higher number of connections than other nodes. The formula for degree variance is
$$\sigma^2 = \langle k \rangle + \frac{\langle k \rangle^2}{r}, \tag{A.2}$$
where r is the dispersion parameter. Degree variance increases with a decrease in r value, leading to heterogeneous connectivity and decreases with an increase in r value, leading to random connectivity similar to the ER networks. The degree distribution for NB networks is given by,
$$p_k = \binom{k+r-1}{k} (1-p)^k p^r, \tag{A.3}$$
where p is success probability defined as $p = r/(\langle k \rangle + r)$.
4. **Scale-free (SF) networks** (see Fig. A.1(d)) follow a power-law degree distribution (Fig. A.2(d)), indicating the presence of nodes with very high degree values, known as "hubs," which lie in the tail of the distribution (Barabási and Bonabeau, 2003). These hubs form the core of SF networks and play a critical role in maintaining network connectivity. Removing a hub can cause significant disconnection in



(a) Erdős-Rényi network (b) Small-world network (c) Neg-Binomial network (d) Scale-free network

Fig. A.1. Illustration of four network structures, generated using the NetworkX library (Hagberg et al., 2008), with $N = 50$ nodes and average degree $\langle k \rangle = 5$: a) ER network, b) SW network with $p = 0.5$, c) NB network with $r = 1$, and d) SF network with $\alpha = 3$.



(a) Erdős-Rényi network (b) Small-world network (c) Neg-Binomial network (d) Scale-free network

Fig. A.2. Degree distribution (histogram) and fitted theoretical degree distributions (black line plots): a) the binomial distribution given by Eq. (A.1) with $N = 50$ and $p = 0.1$ for ER networks, b) a fitted Gaussian distribution with mean 4 and standard deviation 1.2, applied to the degree values from generated SW networks, c) the negative binomial distribution given by Eq. (A.3) with $r = 1.3$, and $p = 0.24$ for NB networks and d) the power-law distribution given by Eq. (A.4) with $\alpha = 3$ for SF networks. Degree distributions are estimated using 100 simulated networks of each type with parameters $N = 50$, $\langle k \rangle = 5$ for the ER network, $p = 0.5$ for the SW network, $r = 1$ for the NB network and $\alpha = 3$ for the SF network, generated using the NetworkX library (Hagberg et al., 2008).

the network, as a disproportionate number of paths traverse through these hubs. An example of an SF network is shown in Fig. A.1(d). Typically, SF networks are generated using the preferential attachment process described in Barabási and Albert (1999), wherein a new node preferentially connects to well-connected nodes. The probability of creating an edge to a node is proportional to its degree. The power-law degree distribution is expressed as:

$$p_k \propto k^{-\alpha}, \tag{A.4}$$

where α is the distribution parameter. The variance of a power-law distribution is given by

$$\sigma = \frac{(\alpha - 1)}{(\alpha - 3)(\alpha - 2)^2}. \tag{A.5}$$

A.2. Sampling algorithms

We compare the RW and the MHRW algorithms for sampling the ER, SW, and SF networks, as described in Appendix A.1.1. Below, we first describe the algorithms:

1. **The RW algorithm** is a classic traversal-based sampling method (Göbel and Jagers, 1974). Sampling starts with selecting an initial node at random from the network. The next node is chosen uniformly from the neighbours of the initial node. These two steps are repeated until the desired sample size is achieved. The probability of choosing a neighbour w of node v is denoted by $P^{RW}_{v,w}$, which is given by

$$P^{RW}_{(v,w)} = \begin{cases} \frac{1}{k_v} & \text{if } w \text{ is neighbour of } v, \\ 0 & \text{otherwise,} \end{cases} \tag{A.6}$$

where k_v is the degree of node v.

2. **The MHRW algorithm** is also a TBS method (Hu and Lau, 2013). Sampling starts by randomly selecting a node from the network. The next node is selected from the neighbours of the initial node using the proposal-acceptance mechanism (Spencer, 2021) based on the degree of the two nodes. Suppose that the last sampled node is v and the proposed next node is w, the probability of accepting w is

$$P^{MH}_{(v,w)} = \begin{cases} \frac{1}{k_v} \min\left(1, \frac{k_v}{k_w}\right) & \text{if } w \text{ is neighbour of } v, \\ 1 - \sum_{y \neq v} P^{MH}_{(v,y)} & \text{if } w = v, \\ 0 & \text{otherwise.} \end{cases} \tag{A.7}$$

Since the RW and MHRW are Markov chain-based algorithms and share the "memoryless" property, some nodes may be revisited during the sampling process, leading to duplicates. The duplicates introduced by revisiting nodes reveal key structural features of the network. For example, they highlight community clustering, as random walks tend to revisit nodes within tightly connected groups (Rosvall and Bergstrom, 2008). Duplicates also emphasise high-degree nodes required for immunisation and intervention strategies (Maiya and Berger-Wolf, 2011). In biological and knowledge graphs, duplicates preserve important clusters and reflect the intensity of interactions, such as gene or drug-target relationships (Dempsey et al., 2012; Zhang et al., 2021).

A.3. Stochastic SIR model

In Appendix A.1, we introduced the structure and properties of the networks; we now demonstrate how to apply a stochastic SIR model on three networks (ER, SW, and SF) to compare disease metric estimates derived from samples obtained using the RW and MHRW algorithms.

Consider a population of size N, which is categorised into three compartments according to the infection status of the individuals during an epidemic. The three compartments are Susceptible (S), Infected (I), and Recovered (R). Let $\boldsymbol{X}(t) \equiv (S(t), I(t), R(t))$ be the state vector where $S(t), I(t), R(t)$ are the number of individuals in state S, I, R respectively at time t. The number of individuals in each compartment of the SIR model changes due to two events: 1) a transmission event and 2) a recovery event, which can be written, respectively, as follows,

$$S + I -> [\beta/N]2I, \tag{A.8}$$

$$I -> [\gamma]R, \tag{A.9}$$

where β/N is the transmission rate per infected individual per unit of time, and γ is the recovery rate per infected individual per unit of time.

The result of these two reactions/events is either the respective increase or decrease in the number of individuals in states $\{S, I, R\}$, which is translated as the change in the state vector $\boldsymbol{X}$. Let $\boldsymbol{v}_i \equiv (v_S, v_I, v_R)$ be the state-change vector, where $i = 1$ denotes reaction (A.8) and $i = 2$ denotes reaction (A.9). The state change vector for reactions (A.8) and (A.9) are $\boldsymbol{v}_1 = (-1, 1, 0)$, and $\boldsymbol{v}_2 = (0, -1, 1)$, respectively.

A.4. Python libraries and functions

Network structures: We use the NetworkX Python library (Hagberg et al., 2008) to generate the ER, SW, NB, and SF networks, described in Appendix A.1.1. ER networks are generated using the `gnp_random_graph` function, SW networks with the `watts_strogatz_graph` function, and NB networks and SF networks with the `configuration_model` function, where the input is a sequence of degree values drawn from a negative-binomial distribution (using the `nbinom.rvs`) and power-law distribution (using the `random.zipf` function respectively from the NumPy library (Harris et al., 2020)).

SIR model simulation: We use the EoN Python library (Miller and Ting, 2020) to implement the Gillespie algorithm (Gillespie, 2007) for simulating a stochastic SIR model (A.3) on a network.

A.5. Mann-Whitney U test hypothesis statements

Hypothesis 1. Null Hypothesis (H_0):

- There is no difference between the probability distribution of the proportion of infected nodes in RW samples and the UN or MHRW samples and the UN.
- The proportion of infected nodes in the UN is greater than or equal to that in the RW or MHRW samples.

Alternative Hypothesis (H_1):
The proportion of infected nodes in the UN is less than that in the RW or MHRW samples.

Hypothesis 2. Null Hypothesis (H_0):

- There is no difference between the distribution of the proportion of infected nodes in RW samples and the UN or MHRW samples and the UN
- The proportion of infected nodes in the UN is less than or equal to that in RW or MHRW samples.

Alternative Hypothesis (H_1):
The proportion of infected nodes in the UN is greater than that in the RW or MHRW samples.

Hypothesis 3. Null Hypothesis (H_0):

- There is no difference between the probability distribution of the number of secondary infections in RW samples and the UN or MHRW samples and the UN
- Number of secondary infections in the UN is greater than or equal to those in RW or MHRW samples.

Alternative Hypothesis (H_1):
The number of secondary infections in the UN is less than that in the RW or MHRW samples.

Hypothesis 4. Null Hypothesis (H_0):

- There is no difference between the distribution of the number of secondary infections in RW samples and the UN or MHRW samples and the UN
- Number of secondary infections in the UN is less than or equal to those in RW or MHRW samples.

Alternative Hypothesis (H_1):
The number of secondary infections in the UN is greater than that in the RW or MHRW samples.

Hypothesis 5. Null Hypothesis (H_0):

- There is no difference between the probability distribution of time-to-infection in RW samples and the UN or MHRW samples and the UN
- Time-to-infection for the UN is greater than or equal to that in RW or MHRW samples.

Alternative Hypothesis (H_1):
Time-to-infection in the UN tends to be less than that in RW or MHRW samples.

Hypothesis 6. Null Hypothesis (H_0):

- There is no difference between the distribution of time-to-infection in RW samples and the UN or MHRW samples and the UN
- Time-to-infection in the UN is less than or equal to that in RW or MHRW samples.

Alternative Hypothesis (H_1):
Time-to-infection in the UN tends to be greater than that in RW or MHRW samples.

Appendix B. Pseudocode for sampling algorithms

We provide the logic of the code for extracting samples using the Random Walk (RW) (Algorithm 1) and Metropolis-Hastings Random Walk (MHRW) (Algorithm 2) algorithms; refer to Appendix A.2 for the mathematical formulation of the algorithms. For both algorithms, to obtain a sample from a graph, we provide three inputs: i) a graph object, which contains the index or label of nodes and the list of connections between nodes; ii) the number of samples, and iii) the sample size, which is the number of nodes required in a sample.

Algorithm 1 Random walk (RW) sampling.

Input: Graph G; Number of samples W; Sample size S
$N \leftarrow$ Number of nodes in G
$E \leftarrow$ Edges in G
Initialize transition matrix $T \leftarrow 0^{N \times N}$
for each edge (i, j) in E **do**
  if $i \neq j$ **then**
    $T[i, j] \leftarrow 1/\text{degree of node}[i]$
    $T[j, i] \leftarrow 1/\text{degree of node}[j]$
  end if
end for
$T\_non\_zero \leftarrow$ matrix of non-zero indices in T
Initialize sample matrix $sample\_array \leftarrow 0^{W \times S}$
$sample\_array[:, 0] \leftarrow$ random sample of W nodes from G as seed nodes
for $itr = 0$ to $S - 1$ **do**
  for each node i in $sample\_array[:, itr]$ **do**
    $next\_node \leftarrow$ randomly select a neighbour of node i from $T\_non\_zero[i]$
    $sample\_array[i, itr + 1] \leftarrow next\_node$
  end for
end for

Appendix C. Statistical test results

C.1. Kolmogorov-Smirnov test

We discuss the normality test results on the data for two disease metrics: proportion of infected nodes and number of secondary infections.

Algorithm 2 Metropolis-Hastings random walk (MHRW) sampling.

Input: Graph G; Number of samples W; Sample size S
$N \leftarrow$ Number of nodes in G
$E \leftarrow$ Edges in G
Initialize $T \leftarrow 0^{N \times N}$
for each edge $(i, j) \in E$ **do**
  if $i \neq j$ **then**
    $T[i, j] \leftarrow \min\left(1, \frac{\text{degree of node}[i]}{\text{degree of node}[j]}\right)$
    $T[j, i] \leftarrow \min\left(1, \frac{\text{degree of node}[j]}{\text{degree of node}[i]}\right)$
  end if
end for
$random\_probs \leftarrow$ matrix of size $W \times S$ with random numbers from 0 to 1
Initialize sample matrix $sample\_array \leftarrow 0^{W \times S}$
$sample\_array[:, 0] \leftarrow$ random sample of W nodes from G as seed nodes
for $itr = 0$ to $S - 1$ **do**
  for each node $i \in sample\_array[:, itr]$ **do**
    $choices[i] \leftarrow$ list of neighbours of node i where $T[i, :] > random\_probs[i, itr]$
    if $choices[i]$ not empty **then**
      $next\_node \leftarrow$ randomly select a node from $choices[i]$
    else
      $next\_node \leftarrow sample\_array[i, itr]$
    end if
    $sample\_array[i, itr + 1] \leftarrow next\_node$
  end for
end for

We perform the normality test with the hypotheses stated in Section 2.1 on the data of disease metrics from the UN, the RW samples, and the MHRW samples. First, we assess whether the data follow a Normal distribution, as it influences the choice of statistical tests for comparing the disease metrics among the RW, MHRW, and UN data. If the data is normally distributed, parametric tests will be used; otherwise, non-parametric tests will be used.

We use the one-sample Kolmogorov-Smirnov (KS) test with a significance level of $\alpha = 0.05$ to assess the normality hypothesis. The KS test statistic measures the largest absolute difference between the cumulative distribution of the data (UN, RW, MHRW) and the normal distribution, which ranges from 0 to 1. A larger KS statistic indicates a greater deviation from the normal distribution. A KS statistics value closer to 1 with a p-value < 0.05 suggests that the data distribution significantly differs from the normal distribution. A value closer to 0 with a p-value > 0.05 implies that the data is significantly similar to a Normal distribution. Below are the results for the two disease metrics across three networks (ER, SW, and SF) for the data from the UN, RW, and MHRW samples.

1. **Proportion Infected Nodes:** For all three networks, the KS statistic (Table C.1) values are approximately 0.5, with p-values < 0.05, evidence to reject the null hypothesis (data follows a Normal distribution) for the UN, RW and MHRW data. The KS statistic is sufficiently large, leading us to conclude that the data distributions for the proportion of infected nodes significantly deviate from a Normal distribution.
2. **Secondary Infections:** For all three networks, the KS statistic (Table C.2) values are approximately 0.5, with p-values < 0.05, evidence to reject the null hypothesis (data follows a Normal distribution) for the UN, RW and MHRW data. The KS statistic is sufficiently large, leading us to conclude that the data distributions for the number of secondary infections significantly deviate from a Normal distribution.

Table C.1
One-sample KS test statistics for the proportion of infected nodes.

Network	UN	RW	MHRW
ER	0.504	0.501	0.501
SW	0.504	0.501	0.501
NB	0.505	0.504	0.500
SF	0.504	0.502	0.507

Table C.2
One-sample KS test statistics for the number of secondary infections.

Network	UN	RW	MHRW
ER	0.578	0.599	0.582
SW	0.509	0.500	0.500
NB	0.650	0.751	0.602
SF	0.509	0.500	0.500

Table C.3
One-sample KS test statistics for the time-to-infection.

Network	UN	RW	MHRW
ER	0.729	0.721	0.726
SW	0.659	0.663	0.662
NB	0.675	0.620	0.661
SF	0.500	0.500	0.500

3. **Time to Infections:** The KS statistic (Table C.3) values are approximately 0.7 for ER, 0.6 for SW, and 0.5 for SF, with p-values < 0.05, evidence to reject the null hypothesis (data follows a Normal distribution) for the UN, RW and MHRW data. The KS statistic is sufficiently large, leading us to conclude that the data distributions for the number of secondary infections significantly deviate from a Normal distribution.

Since data do not follow a Normal distribution across networks, a non-parametric statistical test is required for all three disease metrics to compare the RW and MHRW samples with the UN. We are using the Man-Whitney U-test in this work; results for the same are discussed in Section 3.

Appendix D. Sensitivity analysis - network parameters

We investigate the performance of two sampling algorithms, namely RW and MHRW, in estimating three key disease metrics: the proportion of infected nodes, the average number of secondary infections, and the time-to-infection. This comparison is conducted across four network types: ER, SW, NB, and SF. The study utilises a method outlined in Section 2, where we systematically vary the transmission rate parameter (β) from 0 to 1, alongside other network parameters (as detailed in Table D.1). The systematic variation of parameters allows us to assess how well each sampling algorithm captures the dynamics of infectious disease spread across different network structures, providing insights into their relative accuracy and reliability under varying conditions.

Here, we show the results for all combinations of N and k for the ER networks. For SW networks, a combination of N and p is shown for $k = 4$ only, as for $k \in \{6, 8, 10\}$, the results are qualitatively the same as $k = 4$. For NB networks, results are shown for a combination of N and r for $k = 4$ and $k = 10$, because for $k > 4$, results are qualitatively similar to $k = 10$. For SF networks, results for all combinations of N and α are shown. Please use the code provided on GitHub for obtaining the results not shown here.

Table D.1
List of parameters for sensitivity analysis of four (ER, SW, NB, SF) networks, generated using the NetworkX library (Hagberg et al., 2008).

Network parameter \ Network type	ER network	SW network	NB network	SF network
Number of nodes (N)	{500, 1000, 5000, 10000}	{500, 1000, 5000, 10000}	{500, 1000, 5000, 10000}	{500, 1000, 5000, 10000}
Average degree (k)	{2, 4, 6, 8, 10}	{2, 4, 6, 8, 10}	{2, 4, 6, 8, 10}	–
Edge creation Probability (p)	–	{0, 0.25, 0.5, 0.75, 1.0}	–	–
Dispersion parameter (r)	–	–	{0.5, 1, 2, 5, 10}	–
Power-law exponent (α)	–	–	–	{2, 2.5, 3, 3.5, 4}

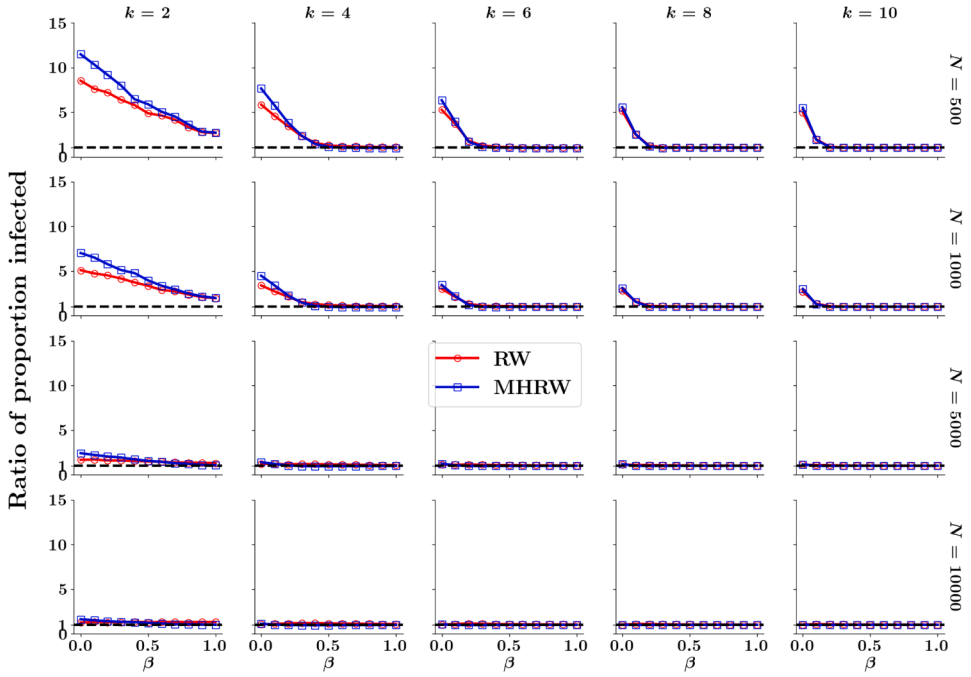


Fig. D.1. Average ratio of proportion infected nodes for ER networks using RW and MHRW sampling algorithms (with duplicates) relative to UN. Analysis considers varying average degree values (k), network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 ER networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

D.1. The effect of network size and average degree

Increasing the network size (N) or the average degree (k) consistently reduces bias in both sampling algorithms across all four network types. In benchmark ER networks, the bias in RW samples decreases by over 50 %, and in MHRW samples by over 25 % as k increases, specifically for $N = 500$ in estimating the proportion of infected nodes (Fig. D.1) and time-to-infection (Fig. D.3). Further, because the bias in the estimates from both sampling algorithms decreases, the percentage reduction in bias also diminishes. For the number of secondary infections (Fig. D.2), MHRW shows an overestimation bias of 1 % to 2 % across N and k values, while RW exhibits a wider range of bias, from 6 % to 19 %. The disease metric, number of secondary infections, is more strongly correlated with the node degree than the other two disease metrics. Consequently, we observe significantly higher bias in the RW estimates than in the MHRW estimates across different values of N and k.

D.2. The effect of clustering

As we increase clustering by decreasing the edge rewiring probability p, transitioning from ER to SW networks, we observe that for $0.25 \leq p \leq 1$, the bias difference between both sampling algorithms is minimal (4 – 5 %) compared to ER networks (> 5 %) for the proportion of infected nodes (Fig. D.4) and time-to-infection (Fig. D.6). This bias difference diminishes further as p decreases, as all nodes become similar in degree, leading to equivalent sampling probabilities for RW (Eq. (A.6)) and MHRW (Eq. (A.7)) algorithms. Like the ER networks, the bias decreases with an increase in N and k. However, in contrast to the ER networks, an increase in clustering reduces bias (up to 10 %) for the number of secondary infections (Fig. D.5) in RW samples, while MHRW maintains a bias within the range of 1 – 2 %. The MHRW sampling algorithm is stable and robust against increased clustering in the network.

In highly clustered networks, specifically lattice networks at $p = 0$, we find that the estimates of three disease metrics align for both sampling algorithms, as all nodes share the same degree, resulting in identical sampling probabilities. However, we observe significantly high bias values compared to $p \geq 0.25$: over 100 % bias for the proportion of infected nodes (Fig. D.4), and 10 – 12 % bias for time-to-infection (Fig. D.6) and the number of secondary infections (Fig. D.5) when $N = 10000$ and $k = 10$. This high bias, even at the maximum values of N and k considered in this study, is attributed to the oversampling of specific regions within the network, limiting the exploration of the entire network due to high clustering and a lack of long-range connections in lattice networks. Unlike the scenario for $p \geq 0.25$, at $p = 0$, increasing N results in higher bias at all k values. Both sampling algorithms exhibit slower mixing, leading to a larger unsampled portion of the network, which increases as N increases.

These findings show that RW and MHRW sampling algorithms do not perform well for highly clustered and structured networks, such as lattice networks. However, the MHRW algorithm effectively reduces size

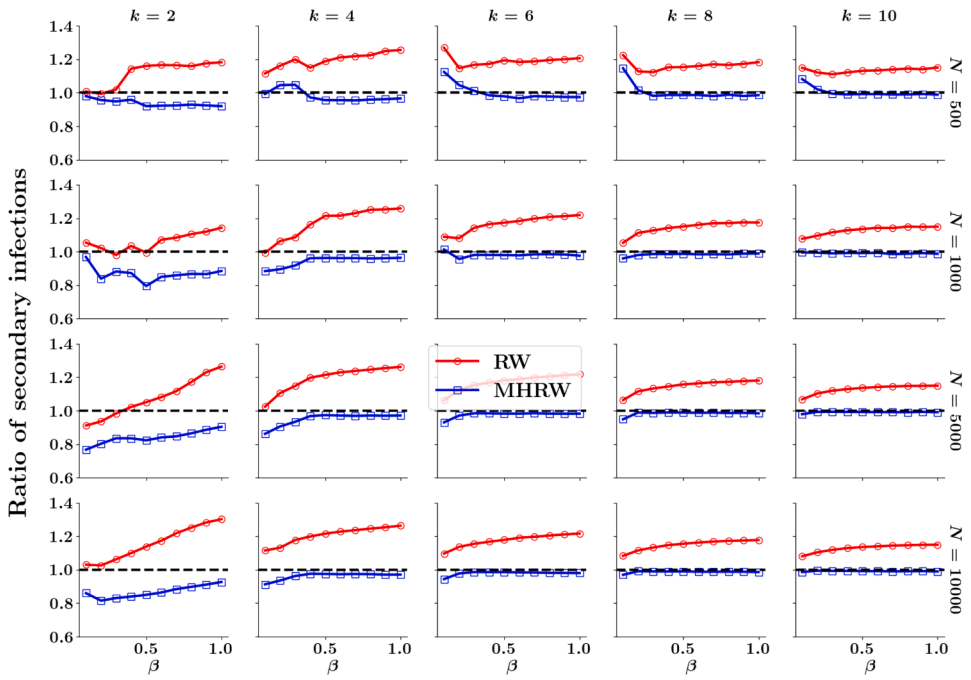


Fig. D.2. Average ratio of number of secondary infections for ER networks using RW and MHRW sampling algorithms (with duplicates) relative to UN. Analysis considers varying average degree values (k), network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 ER networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

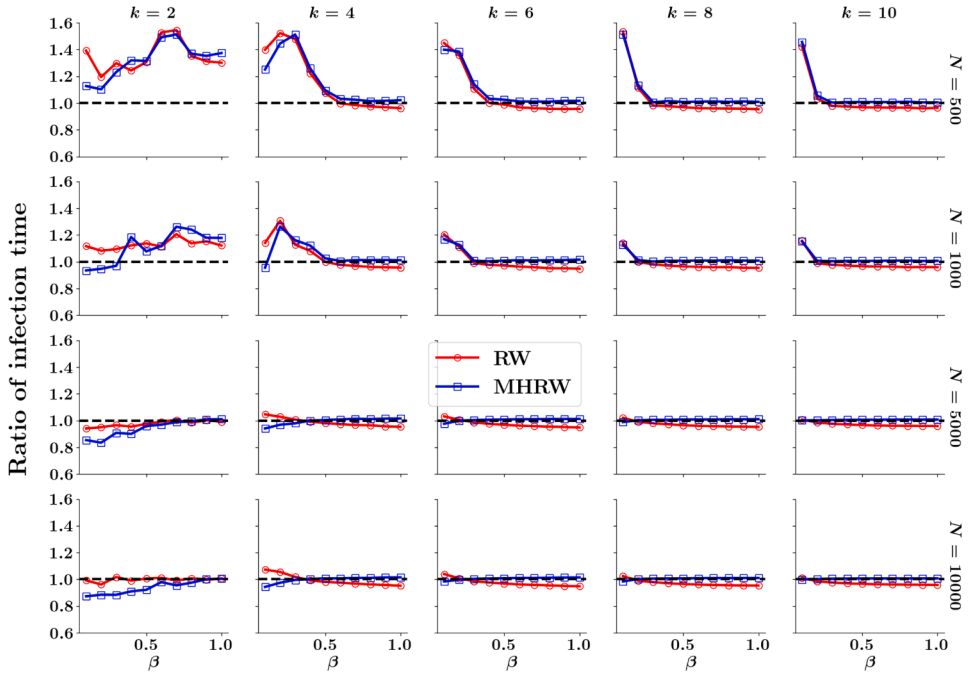


Fig. D.3. Average ratio of time-to-infection for ER networks using RW and MHRW sampling algorithms (with duplicates) relative to UN. Analysis considers varying average degree values (k), network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 ER networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

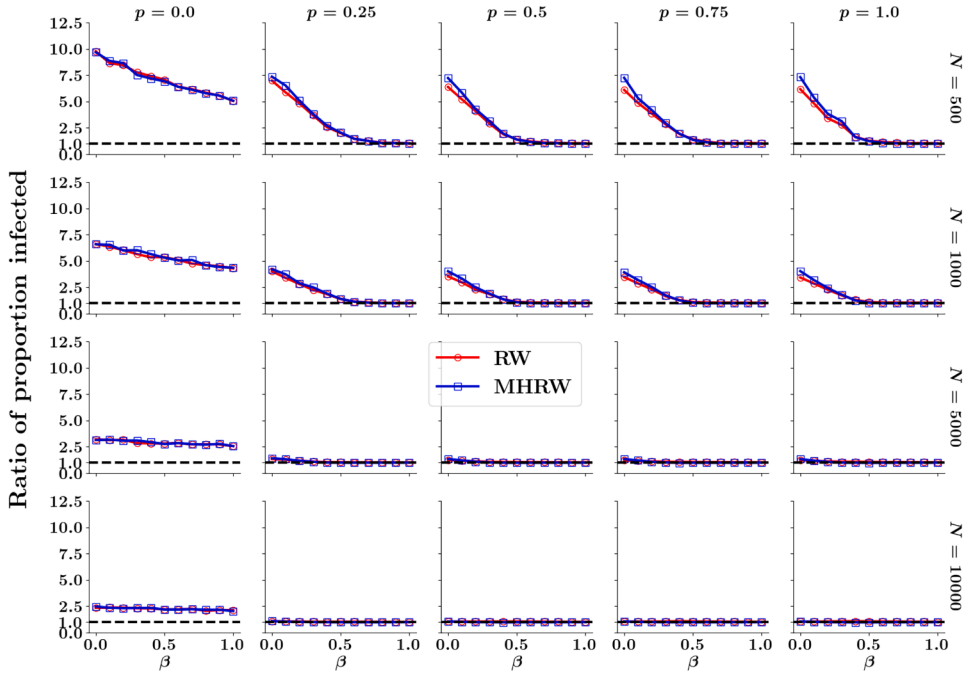


Fig. D.4. Average ratio of proportion infected nodes for SW networks using RW and MHRW sampling algorithms (with duplicates) relative to UN, for $k = 4$, varying edge rewiring probability (p) from 0 to 1, network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 SW networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

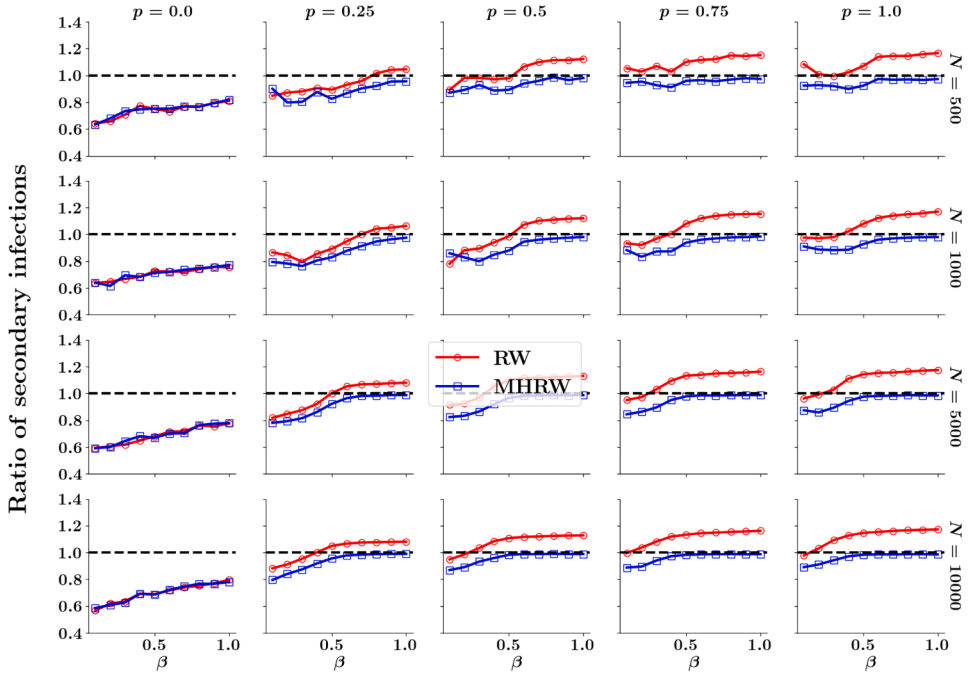


Fig. D.5. Average ratio of number of secondary infections for SW networks using RW and MHRW sampling algorithms (with duplicates) relative to UN, for $k = 4$, varying edge rewiring probability (p) from 0 to 1, network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 SW networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

bias in moderately clustered networks, where the RW algorithm tends to provide biased estimates of disease metrics.

D.3. The effect of degree heterogeneity

Next, we transition from ER networks to NB networks by increasing the variance of the degree distribution. Similar to ER and SW networks (where $p \geq 0.25$), the bias in estimates from both sampling algorithms decreases as the values of N (number of nodes) and k (average degree) increase. However, as we increase degree heterogeneity by reducing the dispersion parameter (r) from 10 to 0.5 in NB networks, resulting in a sixfold increase in degree variance (Eq. (A.2)), we observe a significant amplification of bias in RW sample estimates across the three disease metrics. In contrast, MHRW estimates remain stable and closer to the

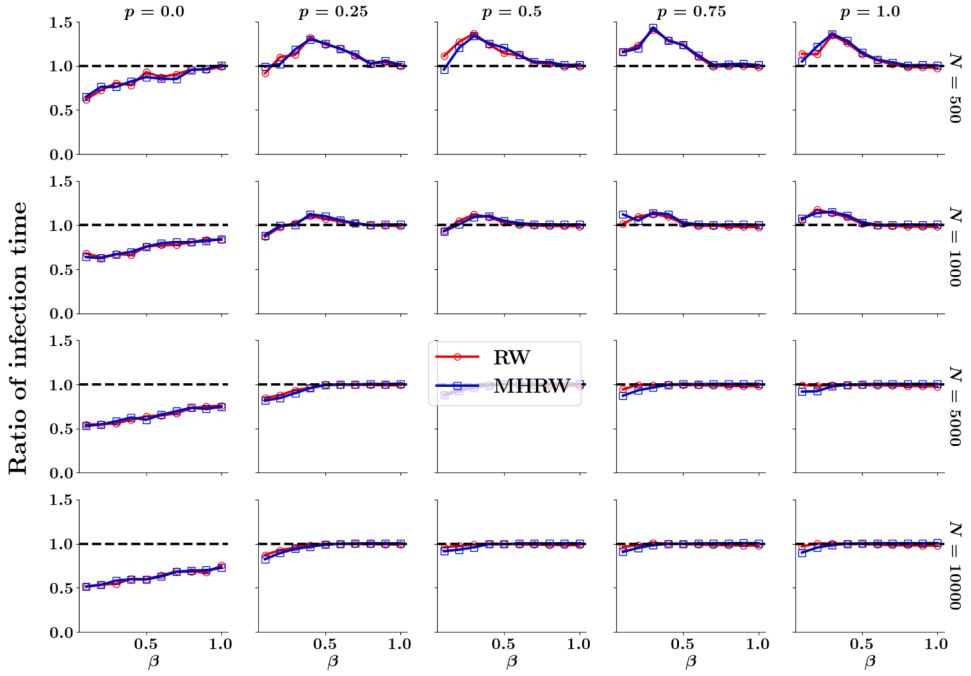


Fig. D.6. Average ratio of time-to-infection for SW networks using RW and MHRW sampling algorithms (with duplicates) relative to UN, for $k = 4$, varying edge rewiring probability (p) from 0 to 1, network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 SW networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

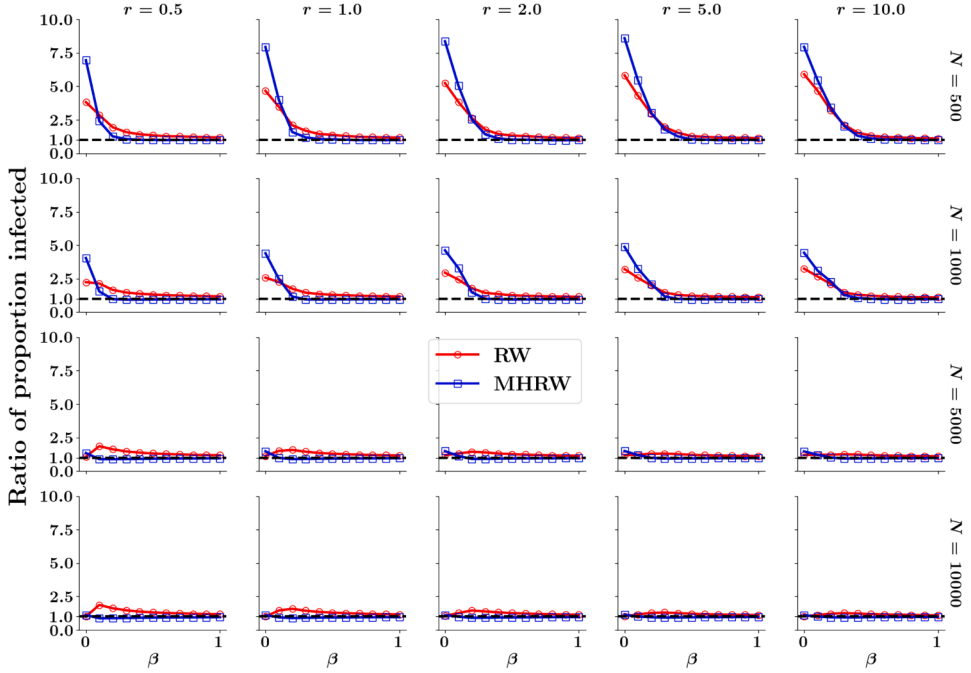


Fig. D.7. Average ratio of proportion infected nodes for NB networks using RW and MHRW sampling algorithms (with duplicates) relative to UN, for $k = 4$, varying dispersion parameters (r), network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 NB networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

underlying network (UN) (see Figs. D.7, D.8, D.12). For instance, at $N = 10000$ and $k = 4$ (Fig. D.7), the bias in the proportion of infected nodes increases from 15 % to 34 % for RW, compared to an increase from 3 % to 6 % for MHRW.

Additionally, the effect of size bias in RW estimates is particularly pronounced for the number of secondary infections in NB networks, due to its strong correlation with node degree. For example, at $N = 10000$ and $k = 4$ (Fig. D.8), RW estimates show a bias of 200 % at $r = 0.5$, while MHRW shows only a 6 % bias.

Regarding time-to-infection (Fig. D.12), we find that for larger values of N and k at all r values, estimates from both algorithms converge towards the UN estimates for NB networks. However, RW tends to have

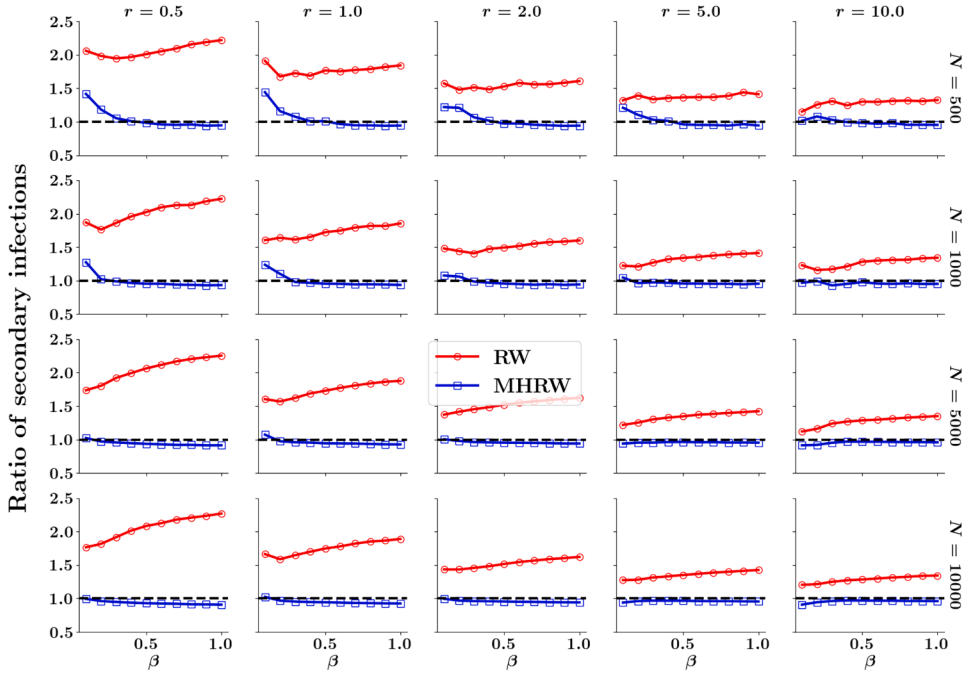


Fig. D.8. Average ratio of number of secondary infections for NB networks using RW and MHRW sampling algorithms (with duplicates) relative to UN, for $k = 4$, varying dispersion parameters (r), network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 NB networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

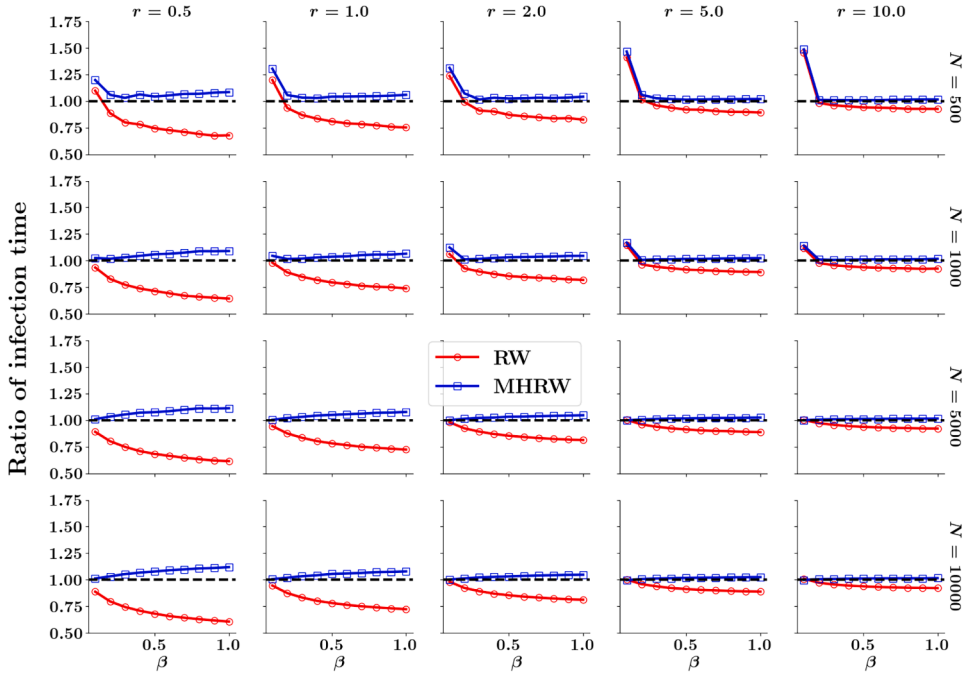


Fig. D.9. Average ratio of time-to-infection for NB networks using RW and MHRW sampling algorithms (with duplicates) relative to UN, for $k = 10$, varying dispersion parameters (r), network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 NB networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

a higher bias. For instance, at $N = 10000$, $k = 10$, and $r = 10$ (Fig. D.9), the bias in RW estimates is 1 %, while MHRW estimates align perfectly with the UN. These observations demonstrate that MHRW significantly reduces size bias for heterogeneous networks.

We evaluate the capability of the MHRW algorithm to reduce size bias by examining degree heterogeneity in extreme cases, specifically for Scale-free (SF) networks, which follow a power-law degree distribution. We will vary the power-law exponent (α) from 2 to 4 to investigate a region of extreme heterogeneity; a decrease in (α) increases the degree of heterogeneity. Importantly, the variance of the power-law distribution (Eq. (A.5)) is infinite for ($\alpha \leq 3$) and finite for ($\alpha > 3$), indicating that we are dealing with sparse and

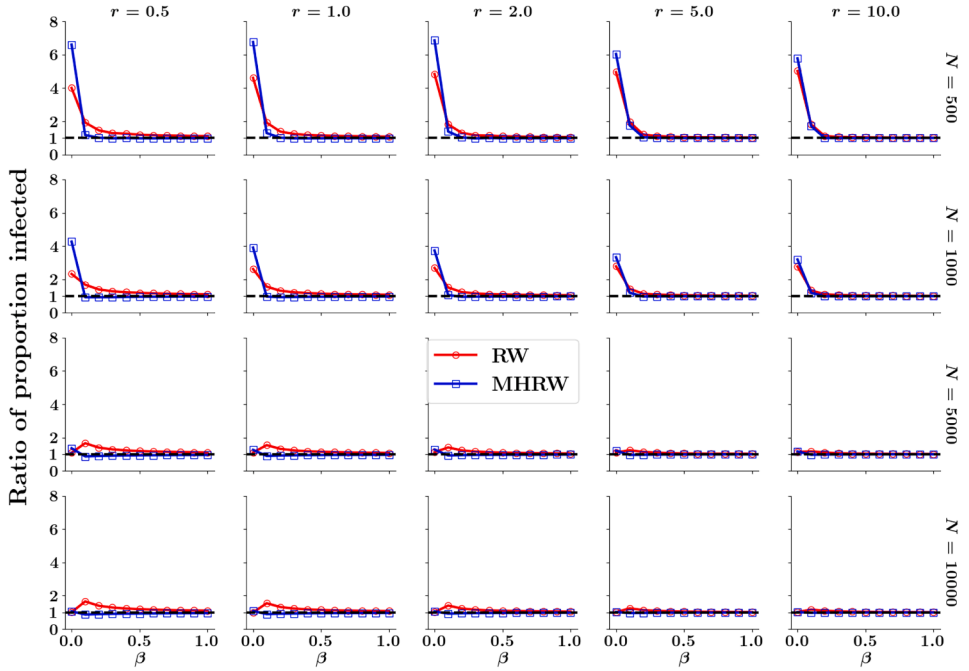


Fig. D.10. Average ratio of proportion infected nodes for NB networks using RW and MHRW sampling algorithms (with duplicates) relative to UN, for $k = 10$, varying dispersion parameters (r), network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 NB networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

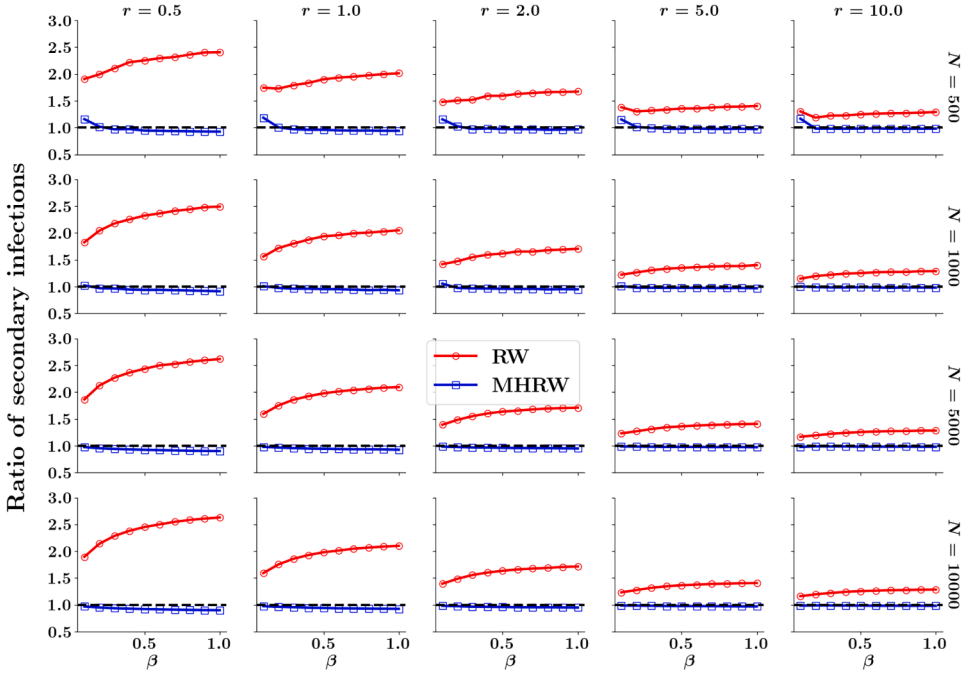


Fig. D.11. Average ratio of number of secondary infections for NB networks using RW and MHRW sampling algorithms (with duplicates) relative to UN, for $k = 10$, varying dispersion parameters (r), network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 NB networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

highly hub-dominated networks compared to NB networks (Figs. D.10, D.11).

In contrast to NB networks, bias is significantly higher for SF networks across both sampling algorithms (Figs. D.13–D.15). Additionally, the pattern of decreasing bias values with increasing N observed in NB networks is absent in SF networks; instead, it is highly dependent on α and the specific disease metric. It is also noteworthy that MHRW does not consistently reduce size bias across all parameter values (N and α) for the three disease metrics. Overall, there is considerable variability in estimates for both sampling algorithms, attributed to the complex network structure, which comprises peripheral and core nodes, having degrees ranging from 1 to 1000 (depending on the value of N).

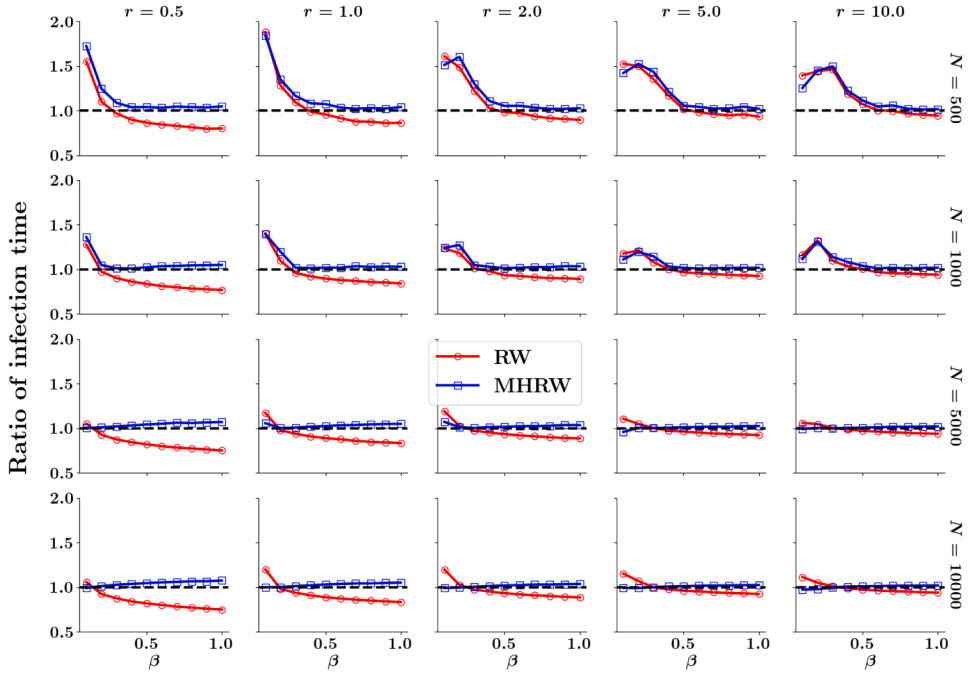


Fig. D.12. Average ratio of time-to-infection for NB networks using RW and MHRW sampling algorithms (with duplicates) relative to UN, for $k = 4$, varying dispersion parameters (r), network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 NB networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

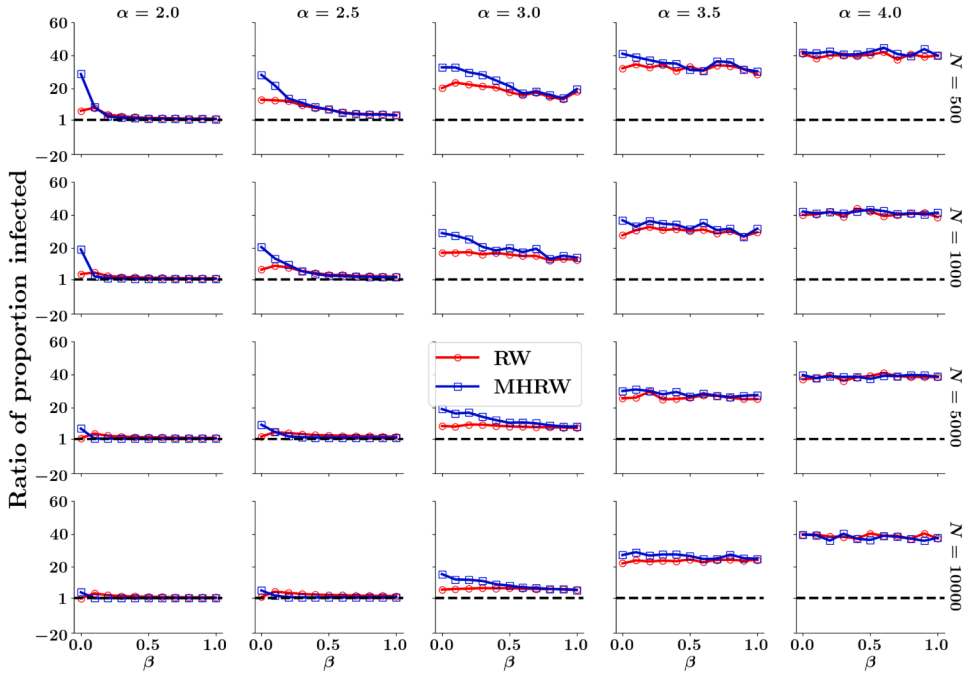


Fig. D.13. Average ratio of proportion infected nodes for SF networks using RW and MHRW sampling algorithms (with duplicates) relative to UN, for varying power-law exponent (α), network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 SF networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

For time-to-infection (Fig. D.15), the bias decreases with increasing N when $\alpha \leq 3$ for both RW and MHRW algorithms, but increases with N for $\alpha > 3$. At extreme heterogeneity ($\alpha = 2$ and $N = 10000$), see bottom left plot in Fig. D.15, RW underestimates time-to-infection by 54 %, while MHRW overestimates it by 7 %, aligning more closely with the UN. Conversely, for the proportion of infected nodes (Fig. D.13), bias decreases with increasing N at all α values for both sampling algorithms. For the number of secondary infections (Fig. D.14), bias increases with N across all α values for both algorithms. For instance, at $\alpha = 3$ (third column from left in Fig. D.14), with an increase in N from 500 to 10000, the overestimation bias by the RW algorithm goes from 986 % to 1500 %, and for MHRW, it goes from 207 % to 290 %.

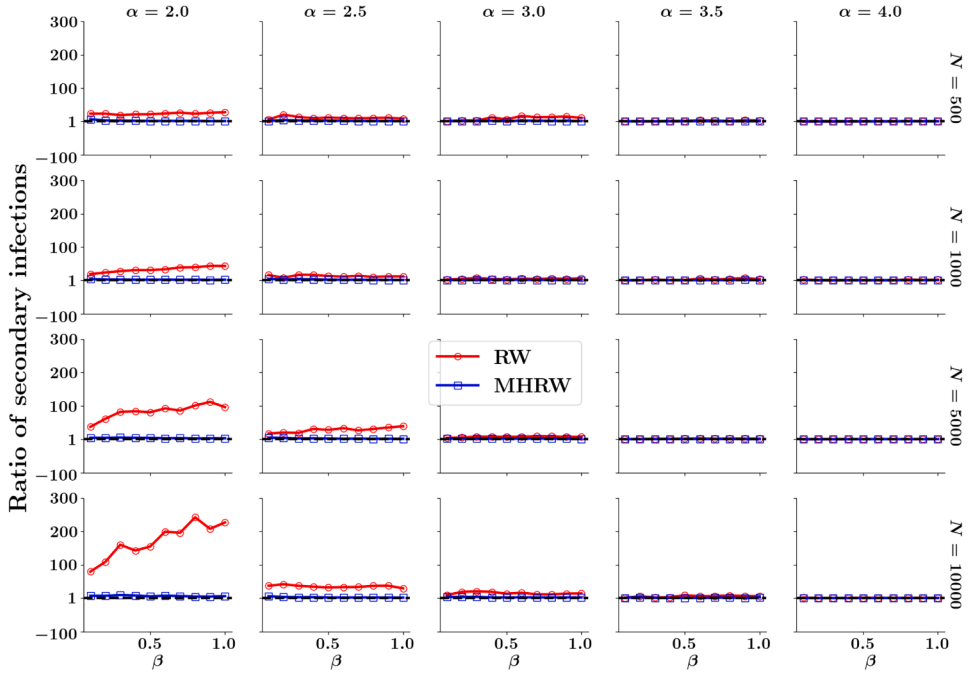


Fig. D.14. Average ratio of number of secondary infections for SF networks using RW and MHRW sampling algorithms (with duplicates) relative to UN, for varying power-law exponent (α), network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 SF networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

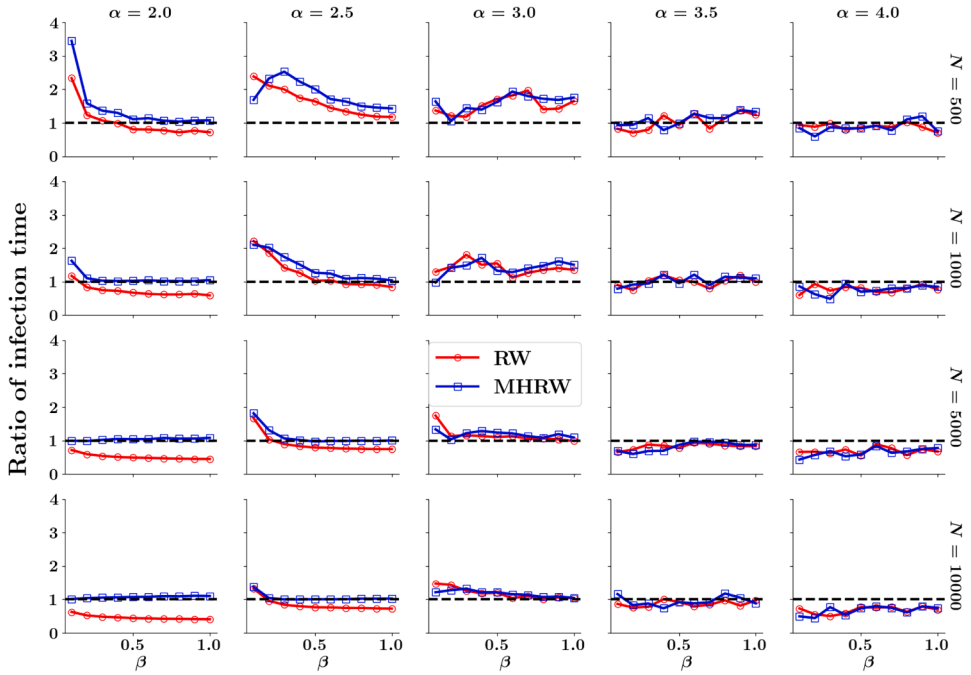


Fig. D.15. Average ratio of time-to-infection for SF networks using RW and MHRW sampling algorithms (with duplicates) relative to UN, for varying power-law exponent (α), network sizes (N), and transmission rates (β) from 0 to 1. UN estimates are based on SIR model simulations for 1000 SF networks (parameters in Table D.1), generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples, each representing 5 % of N, across 1000 networks using RW and MHRW.

These findings indicate the limitation of the MHRW algorithm in reducing size bias for highly heterogeneous networks such as SF networks. Alternative approaches, such as separate sampling of core and peripheral nodes using algorithms like the Sampling for Large-Scale Networks at Low Sampling Rates (SLSR) (Jiao, 2024), may be required for more representative sampling.

D.4. The effect of disease transmission rate

We identified a threshold effect associated with disease transmission rate β for both sampling algorithms, where the bias in disease metric estimates varies depending on whether β is below or above a specific value. When β is below this threshold, both algorithms yield biased es-

timates across all three disease metrics (the proportion of infected nodes, average number of secondary infections, and time-to-infection). In contrast, when β exceeds this threshold, MHRW estimates align closely with the UN estimates, while RW estimates continue to exhibit bias. Furthermore, it was observed that this threshold value decreases as the N and K increase, specifically for ER, SW, and NB networks. Please refer to the figures in Appendix D for a detailed illustration of these relationships.

Appendix E. Disease metrics

We compare the sampling algorithms (RW and MHRW) based on the accuracy of disease metric estimates for the ER, SW and SF networks using the method described in Section 2. Specifically, in Appendix E.1, we compare the sampling algorithms for three disease metrics: 1) proportion of infected nodes, 2) average number of secondary infections, and 3) time-to-infection for the ER and SW networks, with β values from 0 to 1 and $\gamma \in \{1/7, 1/14\}$.

E.1. Sensitivity analysis - recovery rate

In Section 3, we discussed the results obtained with a varying transmission rate β and a fixed recovery rate $\gamma = 1$. Here, we investigate the impact of varying γ and β on the accuracy of disease estimates from sampling algorithms for the ER and SW networks. We compare the disease metric estimates from both sampling algorithms (RW and MHRW) with the UN, using three values of the recovery rate (γ) 1, 1/7, and 1/14. We also discuss the results based on two scenarios: with duplicate nodes in the samples and without duplicate nodes in the samples.

E.1.1. Samples with duplicate nodes

In Fig. E.1, we compare the ratio of the disease metric estimates from the RW and MHRW samples (with duplicate nodes) relative to the UN for three disease metrics: i) proportion of infected nodes, ii) number of secondary infections, and iii) time-to-infection in days. The recovery rate (γ) values are $1, 1/7$, and $1/14$ and transmission rate β increases from 0 to 1. The purple line plots with circle markers are for $\gamma = 1$, the brown line plots with triangle markers are for $\gamma = 1/7$, and the green

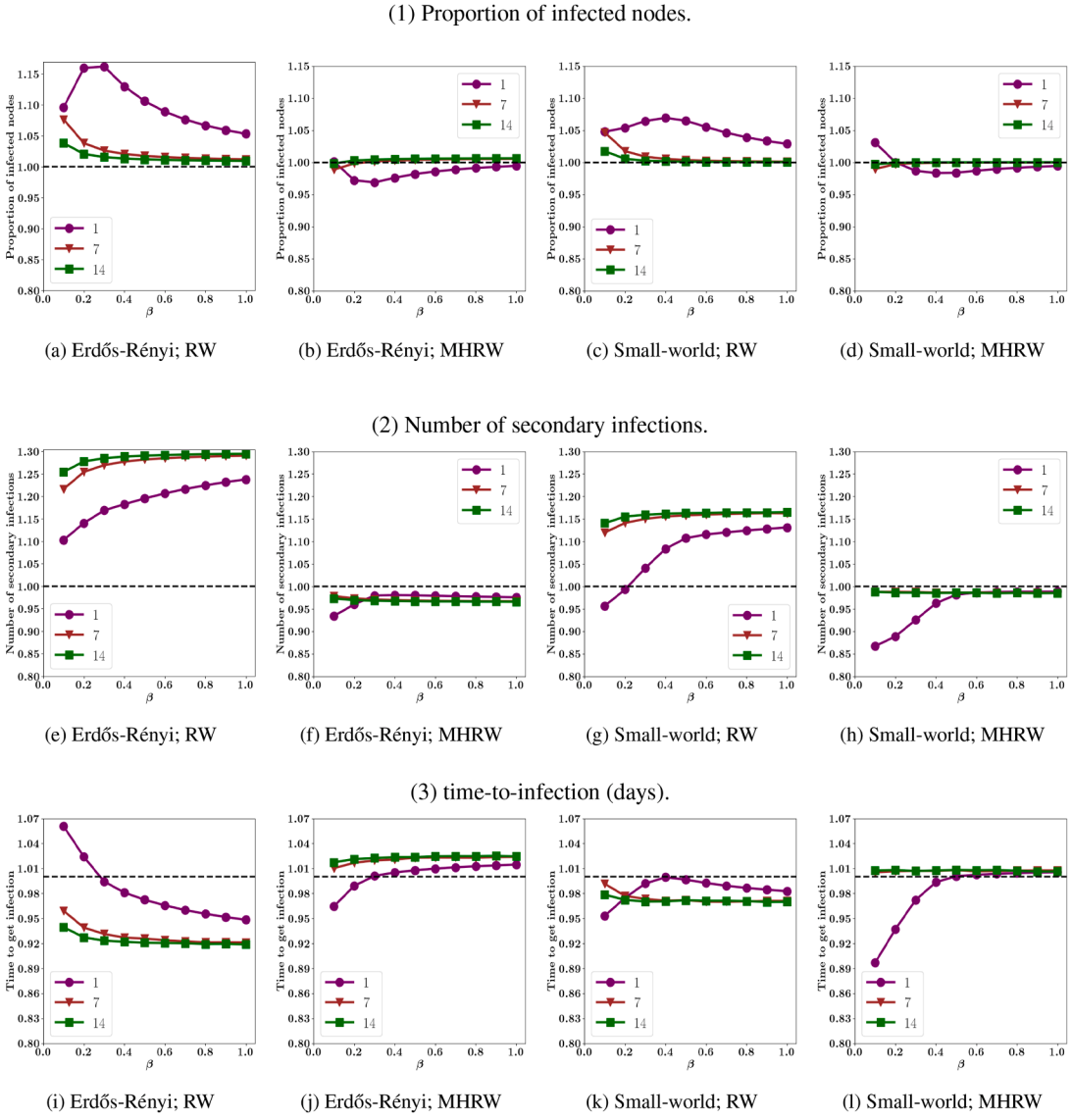


Fig. E.1. The average value of the ratio of three disease metrics estimates for ER and SW networks using RW and MHRW sampling algorithms (with duplicate nodes) relative to the UN; for varying recovery rate $\gamma \in \{1, 1/7, 1/14\}$ and transmission rate β from 0 to 1. The UN estimates are based on SIR model simulations for 10,000 networks of each type (the parameters are found in Table 1) generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples (size 500) for 10,000 networks using RW and MHRW.

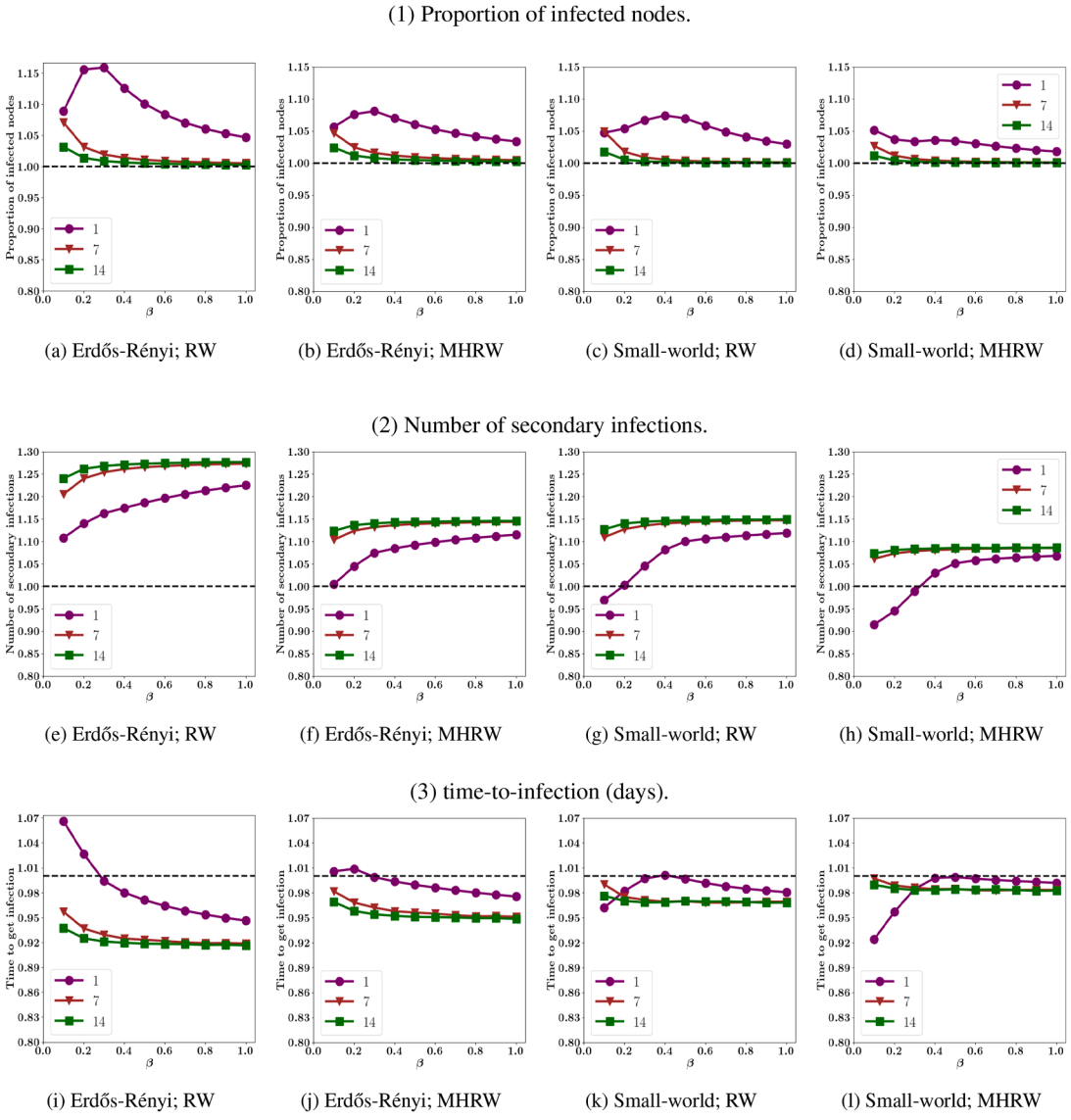


Fig. E.2. The average value of the ratio of three disease metrics estimates for ER and SW networks using RW and MHRW sampling algorithms (without duplicate nodes) relative to the UN; for varying recovery rate $\gamma \in \{1, 1/7, 1/14\}$ and transmission rate β from 0 to 1. The UN estimates are based on SIR model simulations for 10,000 networks of each type (the parameters are found in Table 1) generated with the NetworkX library (Hagberg et al., 2008). Sample estimates are derived from 100 samples (size 500) for 10,000 networks using RW and MHRW.

line plots with square markers are for $\gamma = 1/14$. There are two plots for a network and a disease metric, one for the RW sample estimates and another for the MHRW sample estimates.

For the ER network, as shown in Fig. E.1(a) and (b), the sample estimates for the proportion of infected nodes become closer to the UN as the recovery rate decreases from 1 to 1/14, for both sampling algorithms. Especially, the MHRW sample estimates are well aligned with the UN for $\gamma = 1/14$. For the RW sample estimates, there is a significant drop in overestimation with the decrease in the recovery rate from 1 to 1/14. We observe similar behaviour for the SW networks, as shown in Fig. E.1(c) and (d).

For the ER network, as shown in Fig. E.1(e) and (f), the sample estimates for the number of secondary infections deviate far from the UN as the recovery rate decreases from 1 to 1/14, for both sampling algorithms. The MHRW sample estimates are closer to the UN than the RW sample estimates, irrespective of the recovery rate values. We observe similar behaviour for the SW networks, as shown in Fig. E.1(g) and (h).

For the estimates of the time-to-infection in the ER network, as shown in Fig. E.1(i) and (j), the sample estimates deviate from the UN for both sampling algorithms as the recovery rate decreases. The overestimation increases for the MHRW samples, and the underestimation increases for the RW sample estimates. For the SW networks, as shown in Fig. E.1(k) and (l), underestimation increases for the RW sample estimates. At the same time, the MHRW sample estimates align well with the UN, with a decrease in the recovery rate from 1 to 1/14 for $\beta < 0.4$.

E.1.2. Samples without duplicate nodes

In Fig. E.2, we compare the ratio of the disease metric estimates from the RW and MHRW samples (without duplicate nodes) relative to the UN for three disease metrics: i) proportion of infected nodes, ii) number of secondary infections, and iii) time to get infected in days. The recovery rate (γ) values are 1, 1/7, and 1/14, and transmission rate β increases from 0 to 1. The purple line plots with circle markers are for $\gamma = 1$, the brown line plots with triangle markers are for $\gamma = 1/7$, and the green line plots with square markers are for $\gamma = 1/14$. There are two plots for a network and a disease metric, one for the RW sample estimates and another for the MHRW sample estimates.

For the ER network, as shown in Fig. E.2(a) and (b), the sample estimates for the proportion of infected nodes become closer to the UN as the recovery rate decreases from 1 to 1/14, for both sampling algorithms. For the RW sample estimates, we observe a significant drop in overestimation with the decrease in recovery rate from 1 to 1/14. We observe similar behaviour for the SW networks, as shown in Fig. E.2(c) and (d).

For the ER network, as shown in Fig. E.2(e) and (f), the sample estimates for the number of secondary infections deviate far from the UN as the recovery rate decreases from 1 to 1/14, for both sampling algorithms. We observe similar behaviour for the SW networks, as shown in Fig. E.2(g) and (h).

For the estimates of the time-to-infection in the ER network, as shown in Fig. E.2(i) and (j), we observe that sample estimates deviate from the UN for both sampling algorithms as the recovery rate decreases. The underestimation increases for the RW sample estimates, and the overestimation increases for the MHRW sample estimates. For the SW networks, as shown in Fig. E.2(k) and (l), underestimation increases for the RW sample estimates. At the same time, the MHRW sample estimates align well with the UN, showing a decrease in the recovery rate from 1 to 1/14 for $\beta < 0.4$.

Overall, from comparing the sample estimates for three γ values (1, 1/7, and 1/14), we observe that the estimates for both sampling algorithms are very similar when $\gamma = 1/7$ and $\gamma = 1/14$. However, with $\gamma = 1$, the sample estimates show a noticeable deviation. Our findings suggest that the sample estimates are sensitive to both the recovery rate, the type of disease metric, and the network type.

References

Arratia, R., Goldstein, L., Kochman, F., 2019. Size bias for one and all. Probab. Surv., 16, 1–16.

Balakrishnan, K., 2019. Exponential Distribution: Theory, Methods and Applications. Routledge.

Bansal, S., Read, J., Pourbohloul, B., Meyers, L.A., 2010. The dynamic nature of contact networks in infectious disease epidemiology. J. Biol. Dyn., 4, 478–489.

Barabási, A.L., Bonabeau, E., 2003. Scale-free networks. Sci. Am., 288, 60–69.

Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science, 286, 509–512.

BBC, 2018. Contagion! The BBC four pandemic. https://www.bbc.co.uk/programmes/p059y0p1. BBC Four Television Programme.

Bekara, M.E., Courcoul, A., Benet, J., Durand, B., 2014. Modeling tuberculosis dynamics, detection and control in cattle herds. PloS one, 9, 1–17.

Bishop, C.M., Nasrabadi, N.M., 2006. Pattern Recognition and Machine Learning. Springer.

Bondy, W.H., Zlot, W., 1976. The standard error of the mean and the difference between means for finite populations. Am. Stat., 30, 96–97.

Brauer, F., 2008. Compartmental Models in Epidemiology. Springer Berlin Heidelberg. chapter 2.

Chowell, G., Miller, M.A., Viboud, C., 2008. Seasonal influenza in the United States, France, and Australia: transmission and prospects for control. Epidemiology & Infection, 136, 852–864.

Christley, R.M., Robinson, S.E., Lysons, R., French, N.P., 2005. Network analysis of cattle movement in Great Britain. Proc. Soc. Vet. Epidemiol. Prev. Med.

Conroy, R.M., et al., 2016. The RCSI Sample Size Handbook. A Rough Guide.

Craft, M.E., 2015. Infectious disease transmission and contact networks in wildlife and livestock. Phil. Trans. R. Soc. B Biol. Sci., 370, 20140107.

Craft, M.E., Caillaud, D., 2011. Network models: an underutilized tool in wildlife epidemiology? Interdiscip. Perspect. Infect. Dis., 211, 676949.

Craft, M.E., Volz, E., Packer, C., Meyers, L.A., 2011. Disease transmission in territorial populations: the small-world network of serengeti lions. J. R. Soc. Interface, 8, 776–786.

Cui, Y., Li, X., Li, J., Wang, H., Chen, X., 2022. A survey of sampling method for social media embeddedness relationship. ACM Comput. Surv., 55, 1–39.

Dempsey, K., Duraisamy, K., Bhowmick, S., Ali, H., 2012. The development of parallel adaptive sampling algorithms for analyzing biological networks. In: 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum. IEEE.

Doane, D.P., Seward, L.E., 2011. Measuring skewness: a forgotten statistic? J. Stat. Educ., 19, 18.

Dobos, A., Dobos, v., Kiss, I., 2024. How control and eradication of BVDV at farm level influences the occurrence of calf diseases and antimicrobial usage during the first six months of calf rearing. Ir. Vet. J., 77, 19.

Duncan, A.J., Reeves, A., Gunn, G.J., Humphry, R.W., 2022. Quantifying changes in the British cattle movement network. Prev. Vet. Med., 198, 105524.

Erdös, P., Rényi, A., 1959. On random graphs. Publicationes Mathematicae.

Erdös, P., Rényi, A., et al., 1960. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci., 5, 17–61.

Fielding, H.R., McKinley, T.J., Silk, M.J., Delahay, R.J., McDonald, R.A., 2019. Contact chains of cattle farms in Great Britain. R. Soc. Open Sci., 6, 180719.

Firth, J.A., Hellewell, J., Klepac, P., Kissler, S., group, C. C.-w., Kucharski, A.J., Spurgin, L.G., 2020. Combining fine-scale social contact data with epidemic modelling reveals interactions between contact tracing, quarantine, testing and physical distancing for controlling COVID-19. MedRxiv.

Fournet, J., Barrat, A., 2017. Estimating the epidemic risk using non-uniformly sampled contact data. Sci. Rep., 7, 9975.

Fujimoto, K., Kuo, J., Stott, G., Lewis, R., Chan, H.K., Lyu, L., Veytsel, G., Carr, M., Broussard, T., Short, K., et al., 2023. Beyond scale-free networks: integrating multilayer social networks with molecular clusters in the local spread of COVID-19. Sci. Rep., 13, 21861.

Gani, R., Leach, S., 2001. Transmission potential of smallpox in contemporary populations. Nature, 414, 748–751.

Gillespie, D.T., 2007. Stochastic simulation of chemical kinetics. Annu. Rev. Phys. Chem., 58, 35–55.

Gjoka, M., Kurant, M., Butts, C.T., Markopoulou, A., 2010. Walking in facebook: a case study of unbiased sampling of OSNS. In: 2010 Proceedings IEEE Infocom. IEEE.

Göbel, F., Jagers, A.A., 1974. Random walks on graphs. Stoch. Process. Their Appl., 2, 311–336.

Guerra, F.M., Bolotin, S., Lim, G., Heffernan, J., Deeks, S.L., Li, Y., Crowcroft, N.S., 2017. The basic reproduction number (r0) of measles: a systematic review. The Lancet Infectious Diseases, 17, 420–428.

Hagberg, A., Swart, P.J., Schult, D.A., 2008. Exploring Network Structure, Dynamics, and Function Using NetworkX. Technical Report. Los Alamos National Laboratory (LANL), Los Alamos, NM (United States).

Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E., 2020. Array programming with numpy. Nature, 585, 357–362.

Holme, P., 2013. Extinction times of epidemic outbreaks in networks. PloS one, 8, e84429.

Hu, P., Lau, W.C., 2013. A survey and taxonomy of graph sampling. arXiv preprint arXiv:1308.5865.

Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Hashimoto, Y., Mastrandrea, R., Noël, H., 2011a. Infectious sociopatterns dynamic contact networks. Science Gallery Dublin, Infectious: Stay Away exhibition. Creative Commons Attribution-NonCommercial-ShareAlike license. http://www.sociopatterns.org/datasets/infectious-sociopatterns-dynamic-contact-networks/.

Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J.-F., Van den Broeck, W., 2011b. What's in a crowd? Analysis of face-to-face behavioral networks. J. Theor. Biol., 271, 166–180.

Isoda, N., Sekiguchi, S., Ryu, C., Notsu, K., Kobayashi, M., Hamaguchi, K., Hiono, T., Ushitani, Y., Sakoda, Y., 2025. Serosurvey of bovine viral diarrhea virus in cattle in Southern Japan and estimation of its transmissibility by transient infection in nonvaccinated cattle. Viruses, 17, 61.

Jiao, B., 2024. Sampling unknown large networks restricted by low sampling rates. Sci. Rep., 14, 13340.

Joyal-Desmarais, K., Stojanovic, J., Kennedy, E.B., Enticott, J.C., Boucher, V.G., Vo, H., Košir, U., Lavoie, K.L., Bacon, S.L., 2022. How well do covariates perform when adjusting for sampling bias in online COVID-19 research? Insights from multiverse analyses. Eur. J. Epidemiol., 37, 1233–1250.

Kermack, W.O., McKendrick, A.G., 1927. A contribution to the mathematical theory of epidemics. Proc. R. Soc. London Ser. A, 115, 700-721.

Kiss, I.Z., Miller, J.C., Simon, P.L., 2017a. Mathematics of Epidemics on Networks: From Exact to Approximate Models. Springer International Publishing.

Leskovec, J., Faloutsos, C., 2006. Sampling from large graphs. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Li, R.H., Yu, J.X., Qin, L., Mao, R., Jin, T., 2015. On random walk based graph sampling. In: 2015 IEEE 31st International Conference on Data Engineering. IEEE.

Lieberthal, B., Gardner, A.M., 2021. Connectivity, reproduction number, and mobility interact to determine communities' epidemiological superspreader potential in a metapopulation network. PLoS Comput. Biol., 17, e1008674.

Maiya, A.S., Berger-Wolf, T.Y., 2011. Benefits of bias: towards better characterization of network sampling. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Malmros, J., Masuda, N., Britton, T., 2016. Random walks on directed networks: inference and respondent-driven sampling. J. Off. Stat., 32, 433–459.

Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat., 18, 50–60.

Massey, , F.J.Jr, 1951. The Kolmogorov-Smirnov test for goodness of fit. J. Am. Stat. Assoc., 46, 68–78.

Miller, J.C., Ting, T., 2020. Eon (epidemics on networks): a fast, flexible python package for simulation, analytic approximation, and analysis of epidemics on networks. arXiv preprint arXiv:2001.02436.

Milligan, P., Njie, A., Bennett, S., 2004. Comparison of two cluster sampling methods for health surveys in developing countries. Int. J. Epidemiol., 33, 469–476.

Newman, M., 2018. Networks. Oxford University Press.

Newman, M. E.J., 2006. Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. E–Stat. Nonlinear Soft Matter Phys., 74, 036104.

Newman, M.E.J., Strogatz, S.H., Watts, D.J., 2001. Random graphs with arbitrary degree distributions and their applications. Phys. Rev. E, 64, 026118.

Newman, M.E.J., Watts, D.J., 1999. Renormalization group analysis of the small-world network model. Phys. Lett. A, 263, 341–346.

Nielsen, B.F., Sneppen, K., Simonsen, L., Mathiesen, J., 2020. Heterogeneity is essential for contact tracing. medRxiv .
Nunes, M.C., Thommes, E., Fröhlich, H., Flahault, A., Arino, J., Baguelin, M., Biggerstaff, M., Bizel-Bizellot, G., Borchering, R., Cacciapaglia, G., et al., 2024. Redefining pandemic preparedness: multidisciplinary insights from the CERP modelling workshop in infectious diseases, workshop report. Infectious Disease Modelling, 9, 501–518.
Paton, D.J., Gubbins, S., King, D.P., 2018. Understanding the transmission of foot-and-mouth disease virus at different scales. Curr. Opin. Virol., 28, 85–91.
Qi, M., Tan, S., Chen, P., Duan, X., Lu, X., 2023. Efficient network intervention with sampling information. Chaos Solitons Fractals, 166, 112952.
Richardson, M., Agrawal, R., Domingos, P., 2003. Trust management for the semantic web. In: International Semantic Web Conference. Springer.
Rosvall, M., Bergstrom, C.T., 2008. Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci., 105, 1118–1123.
SeyedAlinaghi, S.A., Abbasian, L., Solduzian, M., Yazdi, N.A., Jafari, F., Adibimehr, A., Farahani, A., Khaneshan, A.S., Alavijeh, P.E., Jahani, Z., et al., 2021. Predictors of the prolonged recovery period in COVID-19 patients: a cross-sectional study. Eur. J. Med. Res., 26, 41.
Spencer, S.E.F., 2021. Accelerating adaptation in the adaptive metropolis–hastings random walk algorithm. Aust. N.Z. J. Stat., 63, 468–484.
Stein, M.L., Steenbergen, J.E.V., Chanyasanha, C., Tipayamongkholgul, M., Buskens, V., Heijden, P.G.M. V.D., Sabaiwan, W., Bengtsson, L., Lu, X., Thorson, A.E., et al., 2014. Online respondent-driven sampling for studying contact patterns relevant for the spread of close-contact pathogens: a pilot study in Thailand. PloS One, 9, e85256.
Suhail, Y., Afzal, J., Kshitiz, 2021. Incorporating and addressing testing bias within estimates of epidemic dynamics for SARS-COV-2. BMC Med. Res. Methodol., 21, 11.
University of Oregon, 2025. Route views project. Accessed: 2025. https://www.routeviews.org/bkup.index.html.
Vynnycky, E., White, R., 2010. An Introduction to Infectious Disease Modelling. OUP Oxford.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. Nature, 393, 440–442.
Wei, W., Erenrich, J., Selman, B., 2004. Towards efficient sampling: exploiting random walk strategies. In: AAAI.
World Organisation for Animal Health, 2024. Foot-and-mouth disease. Accessed: 2024. https://www.woah.org/en/disease/foot-and-mouth-disease/.
Zhang, S., Lin, X., Zhang, X., 2021. Discovering DTI and DDI by knowledge graph with MHRW and improved neural network. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE.