Continuous-Time Simulation of Epidemic Processes on Dynamic Interaction Networks

Rehan Ahmad and Kevin S. Xu

EECS Department, University of Toledo, Toledo, OH 43606, USA
Rehan.Ahmad@rockets.utoledo.edu, Kevin.Xu@utoledo.edu

Abstract. Contagious processes on networks, such as spread of disease through physical proximity or information diffusion over social media, are continuous-time processes that depend upon the pattern of interactions between the individuals in the network. Continuous-time stochastic epidemic models are a natural fit for modeling the dynamics of such processes. However, prior work on such continuous-time models doesn't consider the dynamics of the underlying interaction network which involves addition and removal of edges over time. Instead, researchers have typically simulated these processes using discrete-time approximations, in which one has to trade off between high simulation accuracy and short computation time. In this paper, we incorporate *continuous-time network dynamics* (addition and removal of edges) into continuous-time epidemic simulations. We propose a rejection-sampling based approach coupled with the well-known Gillespie algorithm that enables *exact simulation* of the continuous-time epidemic process. Our proposed approach gives exact results, and the computation time required for simulation is reduced as compared to discrete-time approximations of comparable accuracy.

Keywords: Stochastic epidemic model, SIR model, continuous-time network, dynamic network, rejection sampling, Gillespie algorithm

1 Introduction

Epidemic modeling has been an area of significant interest to the network science community, with applications including the spread of rumors or viral content over social media and the spread of infectious disease over face-to-face contact networks. The distribution and duration of contacts crucially affect the transfer of infection between individuals because it may increase or decrease the chances for infection to occur. Hence, the underlying network topology and properties are very important in epidemic modeling.

Epidemic models are mainly classified into two classes: deterministic and stochastic. We consider stochastic epidemic models, which can be used to simulate a range of possible outcomes for any given set of parameters. The infection process depends upon the instantaneous topology of the network, which is changing continuously over time. Researchers often discretize time into short "snapshots" where the network topology is considered fixed during the time period of a snapshot. This is an approximation of the actual network where its

topology changes only after a fixed interval of time. The epidemic process is then simulated in discrete time over this snapshot-based representation [2, 4]. There is a trade-off between accuracy and computation time that is attached to discrete-time simulation models, depending on the length of the time snapshots.

Continuous-time stochastic epidemic models have traditionally been used for tractability of mathematical analysis [3]. Fennell et al. [6] propose an approach for simulating continuous-time stochastic epidemic models directly using the Gillespie algorithm [8] rather than by using discrete time intervals. They also demonstrated the inaccuracies that can result by using longer intervals for discrete-time simulations. However, the Gillespie algorithm-based approach is not applicable when the network changes over time because the infection rates change based on both the network dynamics (addition and removal of edges) and the infection dynamics. Vestergaard and Génois [12] address this limitation by proposing a method for exact continuous-time simulation on dynamic networks; however, it applies only to discrete-time dynamic networks.

In this paper, we propose an algorithm for simulating continuous-time epidemic processes on *continuous-time dynamic networks* that can change at arbitrary times unlike [12]. Our algorithm combines the Gillespie algorithm-based approach with a *rejection sampling* procedure that rejects inter-event times that occur after a change in the network, i.e. addition or deletion of an edge. We demonstrate that our approach is exact—that is, it correctly simulates the event times in the presence of network changes. We also demonstrate that our rejection sampling Gillespie algorithm results in faster simulations than comparable discrete-time approximations on two real dynamic social network data sets.

2 Background

2.1 Dynamic Interaction Networks

Real networks are generally time-varying in which the edges (interactions) between nodes (individuals) are not fixed or static. Therefore, the underlying network for any dynamic process like infection spreading or information diffusion changes with time, e.g. due to changing patterns of human interactions. The dynamics of this change in network topology may have a non-trivial impact over the processes. Holme [9] discusses several representations used for temporal networks. An exact continuous-time representation includes a sequence of interactions in the form (u, v, t, d), where u and v denote the two nodes involved, t denotes the timestamp of the start of the interaction, and d denotes the duration. The typical discrete-time representation is a sequence of aggregated graphs that represents how the topology of a temporal network changes with time. Each graph in the sequence is an aggregated representation over a time interval.

2.2 Stochastic Epidemic Models

The spread of infectious disease over a population is frequently modeled by a compartmental model in which the population is divided into a set of disjoint

compartments. Some of the most common compartment models are the SIR, SIS, and SEIR models, where S stands for Susceptible, I for Infectious, R for Recovered, and E for Exposed. Any individual in this population exists in one compartment (or group) at a time and assumes similar properties of that group.

In this paper, we consider the SIR model, although our approach generalizes to the other compartmental models as well. In the SIR model, a susceptible individual may get infected after coming in contact with an infectious individual. This individual will recover after a certain amount of time. There are two approaches for modeling the transitions between these groups: deterministic and stochastic. In deterministic models, the transitions between these states are governed by differential equations. In stochastic models, which we consider in this paper, transitions between states occur with certain probabilities [3]. An infectious individual can spread the infection to a susceptible individual with infection probability β. Similarly, an infectious individual can transition to the recovered state with recovery probability μ. Stochastic models are sometimes also simulated to validate analytical results from deterministic models [13].

Discrete-time Models Discrete-time epidemic simulations consider time progressing in constant intervals of length Δt. In a single time interval or snapshot, any individual may make a single transition between compartments. For example, a susceptible individual may transition to the infectious compartment depending upon its contact with other infectious people, or an infectious individual may recover. Each of these transitions happens synchronously at a fixed time. Discrete-time epidemic models are quite convenient for simulation and have been used both with static [1, 10] and dynamic networks evolving over discrete time steps [11], but mathematical analysis is much more difficult in the discrete-time setting, especially for dynamic networks.

Continuous-time Models Despite the simplicity of simulating discrete-time epidemic models, the accuracy and efficiency of discrete-time models is dependent upon the length of time interval being considered. In the continuous-time setting, infection and recovery probabilities are replaced with infection and recovery *rates*, respectively, which denote probabilities per unit time. Allen [3] considers analysis and simulation of continuous-time epidemic models for a fully-connected static network. Fennell et al. [6] consider simulation of continuous-time epidemics over a static (but not necessarily fully connected) network and investigate the effects of discretization of time and its limitations. In both studies, the underlying continuous-time epidemics are simulated using the well-known Gillespie algorithm [8]. Each individual in a population is considered to have an instantaneous rate $r_i(t)$ to transition from one state to another. The Gillespie algorithm works as per the following two properties:

1. The time that the network remains in the same state (no node transitions between compartments) is an exponentially distributed random variable with parameter $\lambda(t) = \sum_i r_i(t)$, the sum of the rates of all nodes in the network.

2. The probability that the next node i to transition from one compartment to another depends on the relative rate of the node $r_i(t)/\lambda(t)$.

Each edge (interaction) with an infectious node (individual) in the network at a certain time increases the infection rate for a susceptible node by an amount β that denotes the rate per infectious neighbor. Similarly, an infectious node will recover at a rate μ. The instantaneous rate r_i for node i to transition between compartments is given by

$$r_i(t) = \begin{cases} \beta m_i(t) & \text{if node } i \text{ is susceptible} \\ \mu & \text{if node } i \text{ is infectious} \end{cases}, \tag{1}$$

where $m_i(t)$ denotes the number of infectious individuals connected to node i [6].

The Gillespie algorithm is not applicable to dynamic networks because $m_i(t)$ can change over time independently from the epidemic process, i.e. when edges with an infectious node are added or deleted. Vestergaard and Génois [12] propose a temporal Gillespie algorithm to simulate continuous-time epidemics over a discrete-time dynamic network. In this paper, we propose a different modification to the Gillespie algorithm to deal with continuous-time dynamic networks.

3 Rejection Sampling Gillespie Algorithm

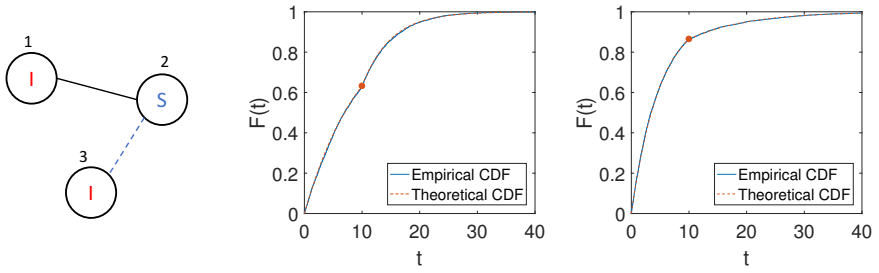
3.1 Inadequacy of Gillespie Algorithm

Fennell et al. [6] propose a method for simulating continuous-time stochastic epidemic models on a static network using the Gillespie algorithm. The exact Gillespie algorithm-based approach is also shown to be much faster than discrete-time approximations that achieve reasonable accuracy. The approach works on a static network because the network topology remains the same over time, hence the number of infectious neighbors an individual can have will remain the same until the next transition (infection or recovery). Thus, the instantaneous transition rate for each node, denoted by $r_i(t)$ in (1), is *constant* until the next transition happens at the simulated event time.

This assumption of static network topology does not hold in case of a dynamic network where edges can be added or removed because the network topology may change before the simulated event time, in which case the simulated event time no longer follows the correct distribution. For example, if a new edge is added with an infectious node i at time t', then $m_i(t')$ in (1) increases by 1, and thus the instantaneous transition rate $r_i(t')$ also increases. Thus, the inter-event time is no longer exponentially distributed as assumed in property 1 of the Gillespie algorithm. Instead, the cumulative distribution function (CDF) of the inter-event time has a knot (instantaneous change in slope) at time t' when $r_i(t')$ increases.

3.2 Theoretical Inter-Event Distribution

The CDF of the inter-event time can be derived analytically as as a continuous function with a series of knots at times when edges are added or removed. With-



(a) Network (b) Edge addition (c) Edge deletion

Fig. 1. Comparison of theoretical CDF and empirical CDF of inter-event distribution for (b) addition and (c) deletion of an edge at time $t = 10$ in 3-node network shown in (a). The red dot in both theoretical CDFs at time $t = 10$ denotes the knot (instantaneous change in slope) when an edge is added or deleted. The two CDFs are almost identical, validating the correctness of our rejection sampling approach.

out loss of generality, we can consider 2 cases: the addition of a single edge at a given time and the removal of a single edge at a given time.

Consider a simple network of three nodes and one edge between node 1 and node 2 as shown in Figure 1(a). Nodes 1 and 3 are initially infectious at time $t = 0$. For a static network where the edge between nodes 2 and 3 is not added, the CDF for the inter-event time, which is the time to infection in this case, is the CDF of an exponential distribution with rate λ given by $P(T \leq t) = 1 - e^{-\lambda t}$, where λ is the infection rate parameter, i.e. the rate at which an infectious individual infects a susceptible individual when they are connected by an edge.

Assume now that, at time t', an edge is being added between an infectious node (node 3) and a susceptible node (node 2). Beginning from the Law of Total Probability and exploiting the memoryless property of the exponential distribution, the CDF for the inter-event time can be shown to be

$$F(t) = P(T \leq t) = \begin{cases} 1 - e^{-\lambda t}, & t \leq t' \\ 1 - e^{-2\lambda t + \lambda t'}, & t > t' \end{cases}$$

The CDF for the case of deletion of an edge between an infectious node and a susceptible node can be derived in a similar manner. Consider again the network in Figure 1(a), but assume now that the edge between nodes 2 and 3 exists at time $t = 0$. At time t', the edge between node 2 and node 3 is deleted. The CDF for the inter-event time is given by

$$F(t) = P(T \leq t) = \begin{cases} 1 - e^{-2\lambda t}, & t \leq t' \\ 1 - e^{-\lambda t - \lambda t'}, & t > t' \end{cases}$$

In the general case where multiple edges are added and removed over time, each addition or removal of an edge with an infectious node creates a new knot in the CDF. This makes it difficult to analytically express the CDF.

Input: Interaction network with timestamps of edge additions and deletions.
1 Initialize starting time as $t = 0$.
2 Randomly select k Infectious nodes to initialize epidemic.
3 Compute transition rate r_i for each Susceptible or Infectious node.
4 Sample the inter-event time for the next transition (infection or recovery) s from an exponential distribution with rate parameter $\lambda = \sum_i r_i$.
5 Update current event time by setting $t = t + s$.
6 **if** $t \leq$ *time of the next added or removed edge* **then** {Sampled time t is valid}
7 Select node i to transition with probability $r_i / \sum_i r_i$.
8 **if** *node i is Susceptible* **then**
9 Change the status of node i to Infectious.
10 **else**
11 Change the status of node i to Recovered.
12 **else** {Sampled time t is invalid}
13 Reject the sampled time t and set $t =$ time of next added or removed edge.
14 If Infectious nodes still exist, go to step 3.

Fig. 2. Exact simulation of continuous-time stochastic epidemic process on a dynamic network using proposed rejection sampling Gillespie algorithm.

3.3 Rejection Sampling Gillespie Algorithm

To overcome the problem with the Gillespie algorithm in a dynamic network setting, we employ the idea of *rejection sampling*[1]. Up to the first knot at time t' in the inter-event time CDF, the CDF matches that of an exponential random variable. Thus, we can sample from the exponential distribution, and if the sampled inter-event time occurs prior to the first knot, then we accept the sample. Otherwise, we reject the sample because the exponential distribution is no longer valid after the knot. We then re-set the current time in our simulation to the time of the knot. Then, the inter-event time will again be exponentially distributed, so we can once again sample the inter-event time from an exponential distribution and decide to accept or reject based on the time of the next knot. We repeat this process until we accept a sample. This approach is valid because the CDF of the inter-event time after each knot is exponentially distributed until the next knot, and the exponential distribution is memoryless. Figure 2 shows the entire algorithm formulated and used for simulation of epidemics in this work.

We evaluate the correctness of our rejection sampling approach on the 3-node network shown in Figure 1(a) by simulating 5000 epidemics for both the edge addition and deletion scenarios and recording the inter-event time for each simulation. Figures 1(b)-(c) show the comparisons between the derived theoretical CDFs from Section 3.2 and the empirical CDFs computed from the simulations. The two plots match almost exactly, confirming the validility of our rejection sampling approach.

[1] Simulating a continuous-time epidemic model using a discrete-time approximation is also sometimes referred to as rejection sampling, e.g. in [6, 12]. We refrain from such terminology in this paper as our proposed rejection sampling approach is exact.

4 Datasets

In this paper, we consider two real-world datasets on face-to-face interactions in a high school setting. The datasets are collected among students of classes in a high school in Marseilles, France using wearable RFID sensors with a proximity range of roughly 1 to 1.5 m [7]. 126 individuals (118 students and 8 teachers) from 3 classes wore the sensors for a period of 5 days in 2011. 180 students from 5 classes wore them for a period of 7 days in 2012. Every 20 seconds, each sensor scanned and recorded the IDs of other sensors in proximity. We convert these 20-second scans to dynamic interaction networks by considering two nodes u and v to have interacted for s seconds if the pair (u, v) shows up in $s/20$ consecutive scans. Since 20 seconds is such a short interval compared to the dynamics of an epidemic process spreading over a dynamic network, these datasets are a good fit for our continuous-time approach, and thus we treat time as varying continuously. The dynamic networks are very sparse—the average instantaneous number of active edges is 5.1 in the 2011 data and 4.0 in the 2012 data.

5 Experiments

To evaluate our proposed rejection sampling Gillespie algorithm, we simulate epidemics on both high school networks using our exact rejection sampling Gillespie algorithm and discrete-time approximations for snapshot lengths $\Delta t \in \{10, 20, 50, 100\}$ seconds. In order to test the robustness of our method, we use a range of values for the infection rate $\beta \in \{0.1, 0.02, 0.002\}$. For each network and each value of infection rate considered, we simulate 1000 epidemics with our exact continuous-time epidemic model and 1000 epidemics with a discrete-time epidemic model for each value of Δt. The recovery rate is fixed at $\mu = 2 \times 10^{-5}$. Each simulation is run until the number of infectious individuals becomes zero, indicating that the epidemic has ended. Figure 3 shows the mean number of Susceptible (S), Infectious (I) and Recovered (R) individuals for continuous- and discrete-time epidemics simulated over the High School 2012 network with infection rate $\beta = 0.1$. Notice that the discrete-time approximations vary significantly in accuracy depending on the snapshot length Δt.

To compare the disease dynamics in the networks we use the area between the normalized mean continuous-time simulation curve and the normalized mean discrete-time simulation curve as our metric, summed over each of the 3 compartments (Susceptible, Infectious, and Recovered). We refer to this as the *normalized error*. This is a variant of an error metric used in [1], with the addition of a normalization step. Since the mean simulation end time for each of the different models may be different, we normalize the computed area with respect to time by dividing it with the minimum mean end time of the models being compared. A lesser value of a normalized area between the plots indicates a better approximation to the simulation outcomes of the exact continuous-time model. The maximum normalized error is 3, consisting of a normalized area of 1 for each compartment. An illustration of the computation of the normalized error metric is shown in Figure 4 for the High School 2012 data with $\beta = 0.1$.

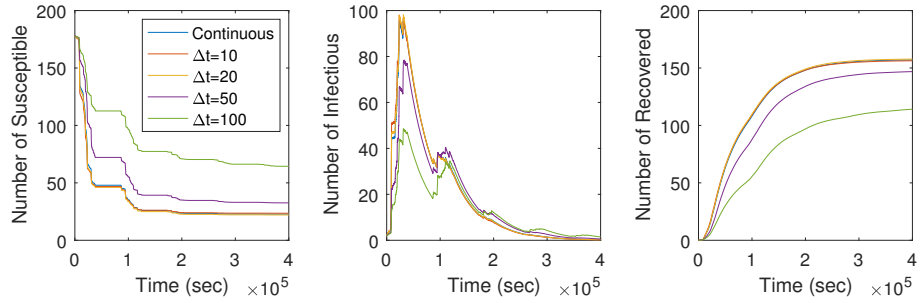


Fig. 3. Mean Susceptible (S), Infectious (I), and Recovered (R) plots over 1000 simulated epidemics for the exact continuous-time model and discrete-time approximations of varying lengths on the High School 2012 dataset with infection rate $\beta = 0.1$. The double peak in the number of infectious over time results from taking the mean of the 1000 simulated epidemics, each with different peak infectious times.

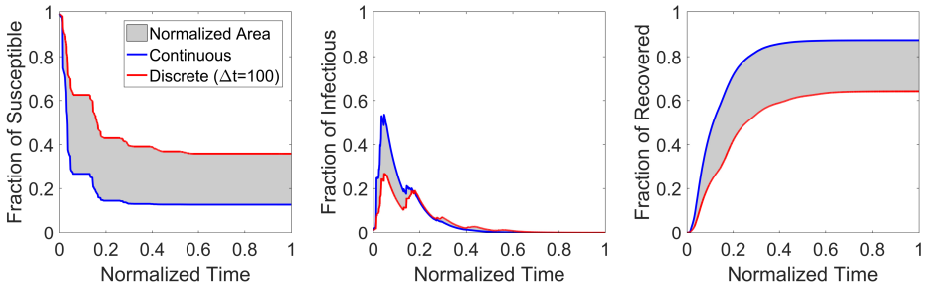
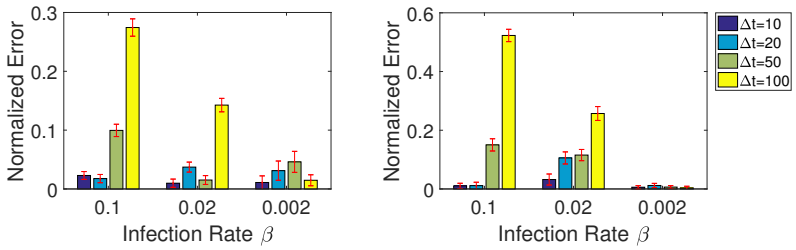


Fig. 4. Computation of normalized error metric by summing the normalized area between the mean curve for simulations with the continuous-time model and the discrete-time model over all three compartments. Smaller normalized error denotes a more accurate discrete-time approximation.

6 Results

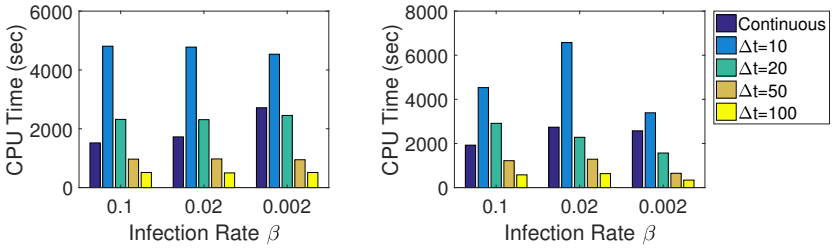
The normalized error for the discrete-time models is shown in Figure 5. To compute the standard error, we use 100 bootstrap replicates [5]. Notice that the overall trend matches what one would expect—as the discrete-time snapshot length increases, the approximation gets worse. This is particularly true for $\Delta t = 100$ seconds, where the error is significantly larger than for smaller snapshot lengths, except for the $\beta = 0.002$ when the infection is not spreading rapidly. We also observe that the error increases as the infection rate β and the snapshot length Δt increase. This observation extends the observation by Fennell et al. [6] in the static network setting to dynamic networks.

The CPU time required to simulate 1000 epidemics using our proposed continuous-time model and each of the discrete-time models is shown in Fig-



(a) High School 2011 (b) High School 2012

Fig. 5. Normalized errors between the mean continuous-time simulated SIR curves and mean discrete-time simulated SIR curves with different Δt values, with error bars denoting standard errors computed using the bootstrap. Error tends to increase with increasing Δt, indicating poorer approximations with longer time snapshots.



(a) High School 2011 (b) High School 2012

Fig. 6. CPU time taken to execute 1000 simulations for each scenario. CPU time for our exact continuous-time simulation is slightly lower than for discrete-time simulation with $\Delta t = 20$. Longer snapshots can be simulated faster but incur higher approximation error as shown in Figure 5.

ure 6. Notice that our proposed continuous-time model is significantly faster than a discrete-time simulation with $\Delta t = 10$ seconds and also slightly faster than with $\Delta t = 20$ seconds in most cases. On the other hand, discrete-time simulations with $\Delta t = 50$ or 100 seconds are faster than our continuous-time simulation, but at the cost of higher approximation error as shown in Figure 5, particularly for $\Delta t = 100$ seconds.

There is a trade-off between computation time and accuracy for discrete-time models. However, since our continuous-time model is exact and faster than discrete-time approximations with extremely short snapshot lengths, we argue that there is no benefit to using discrete-time models with such short lengths. This is because one could use our rejection sampling Gillespie algorithm to simulate the *exact* continuous-time epidemic process in a shorter amount of time! Thus, the only reason to use discrete-time approximations would be due to constraints on computation time, in which case one would use longer snapshots and have to tolerate the loss of accuracy. Otherwise, our proposed continuous-time

approach is superior both in accuracy and computation time and is thus well-suited for general use in simulating epidemic processes over dynamic networks.

Acknowledgements

This material is based upon work supported by the National Science Foundation grant IIS-1755824.

References

1. Ahmad, R., Xu, K.S.: Effects of contact network models on stochastic epidemic simulations. In: Proceedings of the 9th International Conference on Social Informatics. pp. 101–110 (2017)
2. Allen, L.J.: Some discrete-time SI, SIR, and SIS epidemic models. Mathematical Biosciences **124**(1), 83–105 (1994)
3. Allen, L.J.: A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. Infectious Disease Modelling **2**(2), 128–142 (2017)
4. Dong, W., Heller, K., Pentland, A.: Modeling infection with multi-agent dynamics. In: Proceedings of the 5th International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction. pp. 172–179 (2012)
5. Efron, B., Tibshirani, R.J.: An introduction to the bootstrap. CRC Press (1994)
6. Fennell, P.G., Melnik, S., Gleeson, J.P.: Limitations of discrete-time approaches to continuous-time contagion dynamics. Phys. Rev. E **94**, 052125 (2016)
7. Fournet, J., Barrat, A.: Contact patterns among high school students. PLoS ONE **9**(9), e107878 (2014)
8. Gillespie, D.T.: Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry **81**(25), 2340–2361 (1977)
9. Holme, P.: Modern temporal network theory: a colloquium. The European Physical Journal B **88**(9), 234 (2015)
10. Kim, L., Abramson, M., Drakopoulos, K., Kolitz, S., Ozdaglar, A.: Estimating social network structure and propagation dynamics for an infectious disease. In: Proceedings of the 7th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction. pp. 85–93 (2014)
11. Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Colizza, V., Isella, L., Régis, C., Pinton, J.F., Khanafer, N., Van den Broeck, W., Vanhems, P.: Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. BMC Medicine **9**(1), 87 (2011)
12. Vestergaard, C.L., Génois, M.: Temporal Gillespie algorithm: Fast simulation of contagion processes on time-varying networks. PLOS Computational Biology **11**(10), 1–28 (2015)
13. Volz, E.M., Miller, J.C., Galvani, A., Ancel Meyers, L.: Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. PLOS Computational Biology **7**(6), 1–13 (2011)