# SCIENTIFIC REP⚙RTS

**OPEN**

# Simulating SIR processes on networks using weighted shortest paths

Dijana Tolić[1], Kaj-Kolja Kleineberg[2] & Nino Antulov-Fantulin[2]

We present a framework to simulate SIR processes on networks using weighted shortest paths. Our framework maps the SIR dynamics to weights assigned to the edges of the network, which can be done for Markovian and non-Markovian processes alike. The weights represent the propagation time between the adjacent nodes for a particular realization. We simulate the dynamics by constructing an ensemble of such realizations, which can be done by using a Markov Chain Monte Carlo method or by direct sampling. The former provides a runtime advantage when realizations from all possible sources are computed as the weighted shortest paths can be re-calculated more efficiently. We apply our framework to three empirical networks and analyze the expected propagation time between all pairs of nodes. Furthermore, we have employed our framework to perform efficient source detection and to improve strategies for time-critical vaccination.

Modern socio-technological systems can be described as networks with naturally occurring spreading phenomena such as information propagation in social networks[1,2], or the epidemic spread among individuals through their contact network[3–6]. In order to exactly solve the spreading process one must be able to evaluate the probabilistic evolution of the process configurations in time i.e. solve a set of coupled differential equations called master equation[7–10] or Chapman-Kolmogorov equation. However, due to the exponential growth of possible configurations only in small number of special cases[9,10] it is possible to solve the master equation directly. Therefore, different approximations are used: (i) neglecting dynamical correlations by assuming statistical independence at first neighbourhood, pair or node level in mean field approximations[11–16], (ii) neglecting loops in the network structure with message passing or belief propagation techniques[17–19], or (iii) neglecting the temporal evolution by mapping to the percolation process[20–22]. Alternative way is the simulation or sampling the statically correct process configurations or realizations with the dynamic Monte Carlo techniques[23–28].

In this paper, we propose how to exactly transform the SIR spreading dynamics to a shortest path problem on an ensemble of weighted networks using a Monte Carlo approach. In this framework, weights represent interaction time delays along edges. We show the equivalence between the time respecting paths (shortest paths) on the constructed weighted networks and the propagation times of spreading dynamics. The mapping is applicable to the generalized Susceptible Infected Recovered (SIR) model[5] without memory (exponential inter-event distribution/Poisson process) and with memory (arbitrary inter-event distributions). Contrary to the standard dynamic Monte Carlo methods[23–28], we do not have to specify initial conditions upfront and we can sample new realizations from the previous ones by making local random Markov Chain transitions between the weighted networks. In a limit of infinite process time, we establish the connection with bond percolation theory[20,21,29,30]. In the special case of independently assigned edge weights, we establish the connection to the disordered in networks[31–35]. Finally, we show the applications to source detection[18,19,36–39] and time-critical vaccination[40–42].

## Methods

**Mapping SIR dynamics to weighted shortest paths.** Let us start with the Poissonian SIR model, which is memoryless, such that the next state only depends on the present state of the system. This process is defined by two parameters $(\beta, \gamma)$, which describe transitions where susceptible nodes become infected with rate $\beta$ from infected neighbours and transitions where infected nodes recover with rate $\gamma$. This process has exponential inter-event time distributions: $\psi(t) = \beta e^{-\beta t}$ (spread) and $\phi(t) = \gamma e^{-\gamma t}$ (recovery). However, many realistic processes

[1]Laboratory for Machine Learning and Knowledge Representations, Rudjer Bošković Institute, Zagreb, Croatia. [2]Computational Social Science, ETH Zürich, Clausiusstraße 50, 8092, Zürich, Switzerland. Correspondence and requests for materials should be addressed to N.A.-F. (email: anino@ethz.ch)
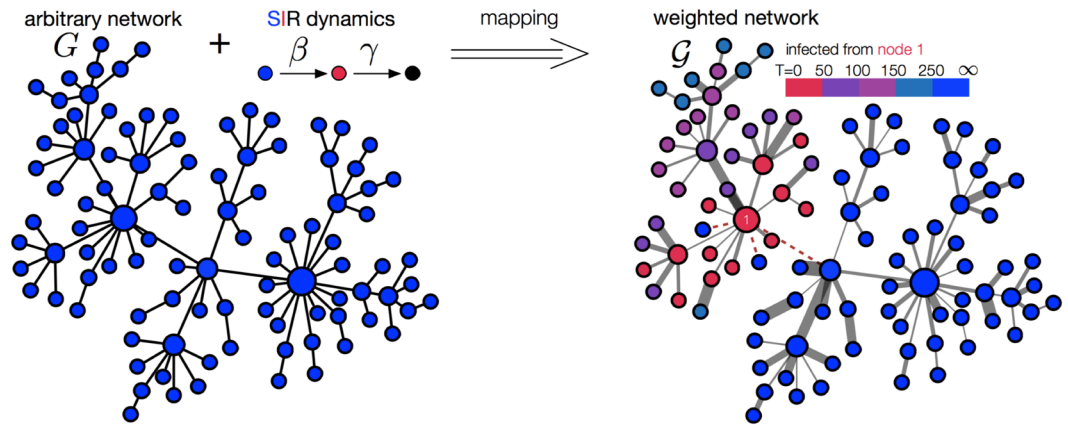
**Figure 1.** The mapping of spreading dynamics to an ensemble of weighted networks. The weights represent the propagation time delays. Each weighted network represents one realization of the stochastic dynamics from an arbitrary source node. The propagation time from the source node to any other node is equal to the shortest path length in weighted network (color coded). The dotted line represents edges with infinite edge weight.

have memory[26] and their inter-event time distributions are non-exponential. Therefore, we allow for arbitrary inter-event density distributions ($\psi(t)$, $\phi(t)$) for generalized SIR processes.

For a given network $G$ and class of generalized SIR spreading models, we show how to create weighted networks $\{\mathcal{G}_k\}$ which encode realizations of the stochastic spreading dynamics. Each weighted network $\mathcal{G}_k$ (see Fig. 1) encodes one possible outcome of the spreading process from every potential source node in a network. In particular, a *time-respecting weighted network* instance $\mathcal{G}_k$ is created by taking the input network $G$ and assigning weights to the edges of the network instance with the Inverse Smirnov transform[43] as

$$\rho_{i,j} = \begin{cases} \Psi^{-1}(x): & \Psi^{-1}(x) \leq \Phi^{-1}(y) \\ \infty: & \Psi^{-1}(x) > \Phi^{-1}(y), \end{cases} \quad (1)$$

where $x$ and $y$ are uniform random numbers $\in [0, 1]$, $\Phi^{-1}(x)$ and $\Psi^{-1}(y)$ are inverse functions of the cumulative inter-event distributions $\Phi(t) = \int_0^t dt' \phi(t')$ and $\Psi(t) = \int_0^t dt' \psi(t')$. The quantities $\Psi^{-1}(y)$ and $\Phi^{-1}(x)$ respectively represent the samples of the transmission and recovery time obtained with the Inverse Smirnov transform of inter-event distributions. In the special case of the Poissonian SIR process one obtains

$$\rho_{i,j} = \begin{cases} -\ln(x)/\beta: & -\ln(x)/\beta \leq -\ln(y)/\gamma \\ \infty: & -\ln(x)/\beta > -\ln(y)/\gamma. \end{cases} \quad (2)$$

If nodes recover faster than a transmission occurs, the weight is set to infinity to indicate no transmission through the edge.

We denote the distance as shortest paths on weighted networks, $d_{\mathcal{G}_k}(v_i, v_j) = \min_{\chi_{ij}} \sum_{(k,l) \in \chi_{ij}} \rho_{k,l}$, where $\chi_{ij}$ is the set of all possible paths from node $v_i$ to node $v_j$ on network $\mathcal{G}_k$ and $\rho_{k,l}$ denotes the weights defined in Eq. (1). Importantly, this distance is equivalent to the propagation time from node $i$ to $j$, i.e. $t(v_i \rightarrow v_j) = d_{\mathcal{G}_k}(v_i, v_j)$ (*time respecting path equivalence*, see Supplementary information, Sec. 1). The run-time complexity of finding shortest paths[44] from specific source node to others is $O(E + N \log N)$, where $N$ denotes the number of nodes and $E$ number of edges in a network. The exact mapping is done by generating a random variable $y$ for each node and variable $x$ for each edge. This takes all dynamical correlations into the account and generates the ensemble of directed weighted networks. A simplified mean-field mapping can be obtained by generating random variables $x$ and $y$ per each edge, independently. This simplified mean-field mapping holds when $\beta \gg \gamma$ (SI case) and generates the ensemble of undirected weighted networks (see Supplementary information, Sec. 5 and 6. for more details). In the section Additional information of this paper, we provide a link to the code that implements the proposed mapping.

**Sampling methods for simulation.** The instances of the ensemble of weighted networks can be sampled either (i) independently, or (ii) by traversing elements of the ensemble with Markov Chain.

The Markov Chain (rejection-free Gibbs sampler – see Supplementary information, Sec. 3) consists of transitions $W_{i,j} = w(\mathcal{G}_i \rightarrow \mathcal{G}_j)$ over the set of weighted networks $\{\mathcal{G}_k\}$, where each weighted network corresponds to one realization of the spreading process from every potential source node. This allows us to construct transitions between weighted networks by changing the weights in the first neighbourhood of a randomly selected node for exact mapping or by assigning a new weight to the randomly selected edge for a mean-field mapping, both according to Eq. (1) (see Fig. 2). The existence of a stationary distribution of this Markov Chain is guaranteed by detailed balance property and the uniqueness by ergodicity (see Supplementary information, Sec. 3). In other words, the Gibbs sampling—as well as independent sampling—constructs the probability distribution $P(\mathcal{G})$ of the
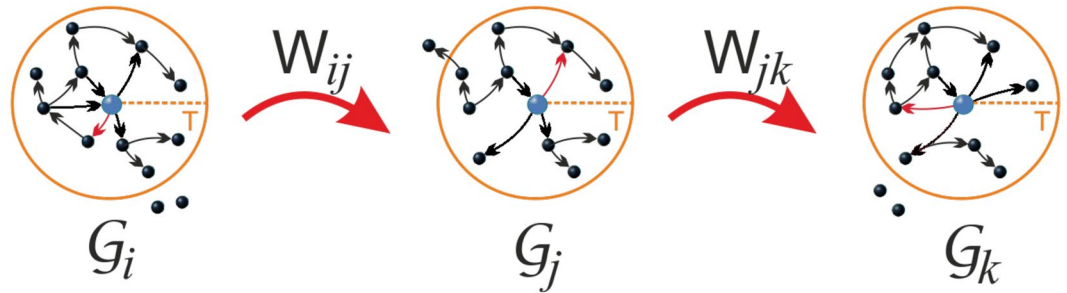
**Figure 2.** Markov Chain Monte Carlo sampling. The transitions $W_{i,j} = w(G_i \rightarrow G_j)$ between different instances of weighted networks in the ensemble. Randomization of the weight of an edge (denoted by red) corresponds to the transition between the states (realizations of the process).
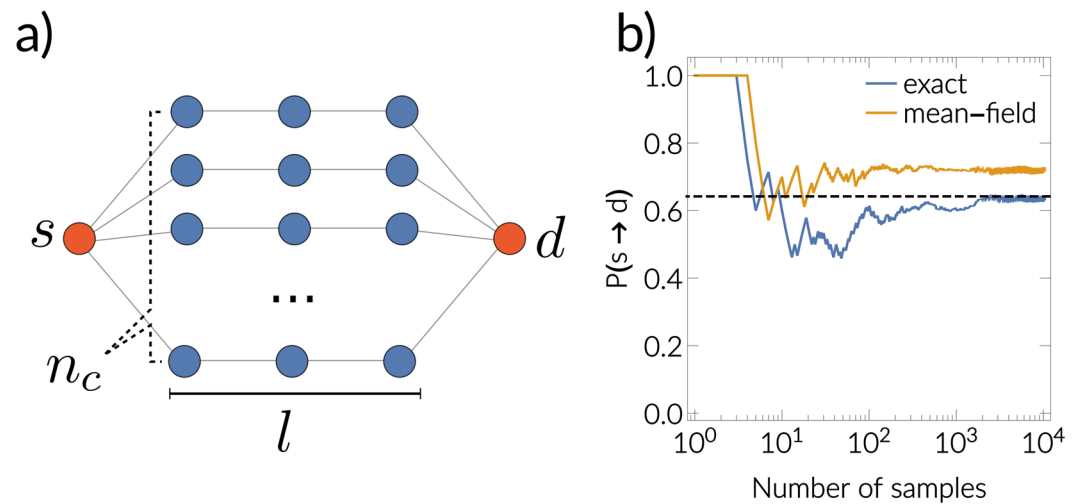


**Figure 3.** Toy network model example. (**a**) Sketch of simple toy network, that consists of $n_c$ chains with $l$ nodes. The probability that source node $s$ infects destination node $d$ can be calculated analytically (see supplementary information, Sec. 5. for details) and is given by $P(s \rightarrow d) = 1 - \sum_{j=0}^{n_c} p_{n,j} \cdot \left(1 - p_{1,1}^l\right)^j$. The term $p_{n,k}$, denotes transmissibility for the first neighborhood Eq. 10 for Poissionan SIR process and Eq. 9 for generalized non-Markovian process. (**b**) Simulation results on the toy network with $n = 20$, $l = 3$ with Poisson SIR ($\beta = 1$, $\gamma = 1$). The blue curve represents the estimations with the exact mapping (conditional independence among edge weights), the red curve represents the mean-field mapping (weights are independent) and the black dotted line represents the analytical solution (see supplementary information, Sec. 5. for a million size toy network experiment).

weighted network ensemble, such that it is a statistically exact representation of the stochastic process. From the ensemble of weighted networks, due to ergodicity, the expectation $\langle f(\mathcal{G}) \rangle$ can be estimated as the average over $n$ samples of weighted networks:

$$\langle f(\mathcal{G}) \rangle = \sum_{\mathcal{G}_k} P(\mathcal{G}_k) f(\mathcal{G}_k) \approx 1/n \sum_{k=1}^{n} f(\mathcal{G}_k) \tag{3}$$

By changing functions $f(\mathcal{G})$, we estimate different properties such as total outbreak size, temporal evolution, expected propagation time or source likelihood (see Supplementary information, Sec. 4). The same estimation formula 3 can be applied to the independent sampling. The convergence rate of independent estimation are bounded with the Berry–Esseen inequality as $O(n^{-1/2})$. If all to all shortest paths need to be calculated, after each weight change in Markov Chain transition, shortest paths can be dynamically recalculated[45,46] with computational complexity $O(N^2 log^3 N)$, where $N$ denotes the number of nodes in a network (see Supplementary information, Sec. 8 for more details). In Fig. 3, we demonstrate that the estimations converge to the analytical solution on a toy network model. This confirms that the shortest paths in the ensemble of weighted networks are taking into the account stochastic spreading along all possible paths in the original unweighted network.

**Relation to bond percolation.** In the following, we state the connection to percolation theory. From the perspective of source node $v_i$, the shortest path $d_{\mathcal{G}}(v_i, v_j)$ indicates the first infection time of node $v_j$. Therefore all
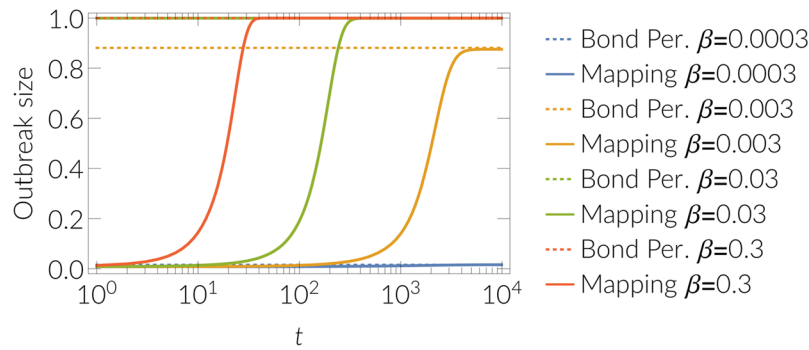
**Figure 4.** Equivalence to bond percolation. Average outbreak size in time estimated with the proposed mapping and comparison with bond percolation late-time outbreak size (dotted lines), see Eq. 4. In a limit of infinite time the mapping becomes equivalent to the bond percolation. Results were obtained for the SIR dynamics (continuous time) with different transmission rates $\beta$ (0.3, 0.03, 0.003, 0.0003) and recovery rate $\gamma = 0.001$ and with $n = 10^4$ simulations on the 4-connected two dimensional regular lattice ($11 \times 11$).

nodes that are not reachable within the temporal distance $t$ stay in the susceptible state. If a node is reachable within the temporal distance $t$, then it was infected. By letting $t$ go to infinity we obtain asymptotic realization *equivalence to bond percolation*. In particular, the size of the set of nodes $v_j$ that are reachable in any finite time from the source node $v_i$, $|S_p(v_i)|$, is given by

$$|S_p(v_i)| = \lim_{t \to \infty} |\{d(v_i, v_j) < t\}|. \tag{4}$$

The full proof is given in the Supplementary information, Sec. 2. and holds for simplified mean-field mapping, where edge weights are independent.

Hence, $|S_p(v_i)|$ denotes the size of the connected bond percolation component, where the transmissibility parameter[29] from source $v_i$ is given by

$$p = \int_0^\infty d\tau \phi_i(\tau) \int_0^\tau d\tau' \psi_i(\tau'). \tag{5}$$

In Fig. 4, we present the experiments where the equivalence to bond percolation is validated. The transmissibility quantifies the probability that an infected node transmits infection along a link to a susceptible neighbour before it recovers. The second integral accounts for the conditional probability of transmission up to the fixed recovery time, and the first integral account for all possible recovery times $\tau$. Furthermore, to take dynamical correlations into the account that were previously missing in[29], we generalize and formalize the transmissibility for the first neighborhood or, more precisely, the probability $p_{n,k}$ that $k$ out of $n$ directed links will be active

$$p_{n,k} = \int_0^\infty \phi(\tau) d\tau \binom{n}{k} (1 - \Psi(\tau))^{(n-k)} \Psi(\tau)^k, \tag{6}$$

where $\Psi(\tau) = \int_0^\tau \psi(t) dt$. For Poissionan process this becomes

$$p_{n,k} = \binom{n}{k} \frac{\gamma}{\beta} \frac{\Gamma(k+1)\Gamma\left(\frac{\gamma + \beta(n-k)}{\beta}\right)}{\Gamma\left(k + 1 + \frac{\gamma + \beta(n-k)}{\beta}\right)}. \tag{7}$$

The generalized transmissibility for the first neighborhood Eq. 7 establishes the connection to semi-directed bond percolation[30] (see Supplementary information, Sec. 5 and 6). To have exact mapping of the SIR process to percolation, having a single percolation parameter $p$ is no longer accurate when $\beta \approx \gamma$ and generalized transmissibility $p_{n,k}$ for the first neighborhood is needed to capture all dynamical correlations.

**Relation to disordered networks.** The question how the average propagation time between two individuals scales with the system size is crucial for many real applications and of special theoretic interest. In particular, if this distance scales logarithmically with the system size, an epidemic can spread nearly instantaneously in arbitrary large networks. However, if it scales like $N^\sigma$, in large enough systems the disease will not spread through the network in realistic timescales. We can address this question for the SI-model ($\gamma = 0$) in an elegant fashion by establishing an analogy to the shortest path problem in disordered networks[31–35], which are generated by assigning weights $\rho$ independently to edges drawn from some distribution $f(\rho)$. In[31] the authors have shown that the shortest path length from node $v_i$ to node $v_j$ in these weighted networks scales differently with the system size depending on $f(\rho)$, and distinguish between strong and weak disorder. The shortest path length is defined as $d(v_i, v_j) = \min_{\chi_{ij}} \sum_{(k,l) \in \chi_{ij}} \rho_{k,l}$, where $\chi_{ij}$ is the set of all possible paths. For the example of Erdos-Renyi networks, in the case of strong disorder the shortest path length scales as $\langle d \rangle \propto N^{1/3}$, and for weak disorder as $\langle d \rangle \propto \log N$. The temporal distance we have defined earlier is equivalent to the shortest paths on weighted graphs. Therefore,
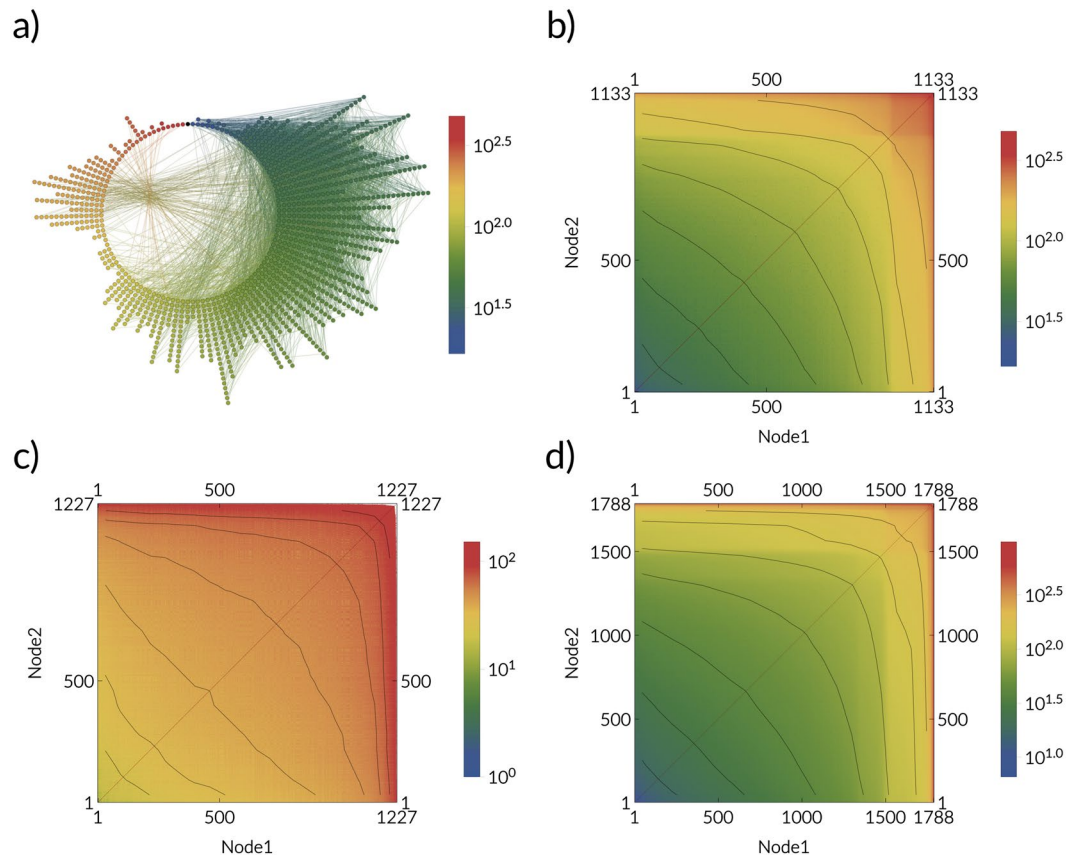
a)



b)



c)



d)



**Figure 5.** The expected propagation time $D_{ij}$ on empirical networks. The average propagation time is calculated as the average shortest path distances in the ensemble of weighted networks (mapped dynamics). **(a)** Nodes in the email network network are placed on a circle, where the angular coordinate represents the average propagation time from the source node (black) with index 500, increasing in the clockwise direction (from blue to red). The average propagation time $D_{ij}$ for: **(b)** the email network[48] ($\beta = 0.01$), **(c)** the airport network ($\beta$ estimated from flux data see Supplementary information, Sec. 7), **(d)** the Petster network[49] ($\beta = 0.01$). Nodes are ordered by their characteristic spreading timescale $\tau_i$, see Eq. 8) with condition $\overline{N} = N/2$.

we can use this analogy to study the scaling of the average propagation time with the system size. From[31] it is know that if $f(\rho)$ is taken as the exponential distribution, only weak disorder can occur, and the average distance scales with the logarithm of the systems size for Erdos-Renyi networks. This means that for the memory-less spreading process, where $\psi(\cdot)$ is the exponential distribution, the average propagation time scales with the logarithm of the system size and diseases spread quickly in systems of any size. However, if the inter-event distribution is for instance given by a lognormal distribution (non-Markovian case), we know from[31] that strong disorder can occur such that the average distance and hence the average propagation time scales like $\langle d \rangle \propto N^{1/3}$. In this case, the finite time properties of the spreading process are drastically different to the non-Markovian case, as the average propagation time becomes very large as the size of the system increases, such that the disease will not spread globally in realistic timescales. To conclude, although it has recently been shown that the infinite time steady states of non-Markovian processes can be mapped to Markovian ones[47], their finite time dynamics can be dramatically different.

**Dataset availability.** All data that we used in our analysis are freely available[48–50].

**Main code available at.** https://github.com/ninoaf/SpreadingMapping

## Results

**Expected propagation time between node pairs.** Starting from a given node, what is the expected time after which the disease reaches each other node in the network? The answer to this question is encoded in the ensemble of weighted networks. For SI processes ($\gamma = 0$), the expected propagation time is always finite and can be estimated as the average over $n$ samples of weighted networks as $D_{ij} = 1/n \sum_{k=1}^{n} f(\mathcal{G}_k)$, where $f(\mathcal{G}_k) = d_{\mathcal{G}_k}(v_i, v_j)$ denotes the length of weighted shortest path between $v_i$ and $v_j$. Note, that for the SIR process, the expected propagation time is not finite i.e. it diverges. In Fig. 5(a), we show an explicit example from the perspective of a specific source node. The advantage of the temporal maps we have created is they encode this type of information between all possible node pairs. The result is shown in Fig. 5(b–d) for three empirical networks: the
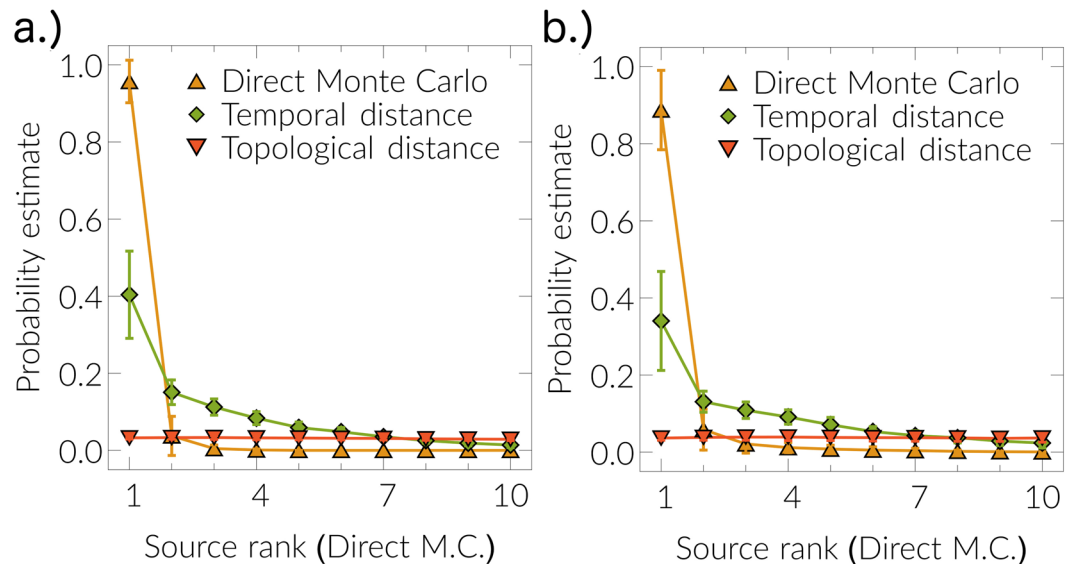
**Figure 6.** Source detection. We compare source detection performance for different methods for the SIR model: (**a**) $\beta = 0.7$, $\gamma = 0.3$ and (**b**) $\beta = 0.7$, $\gamma = 0.7$ at time $T = 5$ (discrete time) on the 4-connected lattice ($30 \times 30$ nodes). For each observed realization, we rank nodes according to their estimated probability score according to the Direct Monte Carlo method[36], in case of a tie we prioritize the node with a higher estimator according to the temporal distance method. We compare different methods as explained in the text: Direct Monte Carlo source probability ranking, source detection ranking by topological distance[54], and the estimation from our mapping (temporal distance). Results are averaged over 30 observed realizations and error bars show one standard deviation, from top to bottom.

email network[48], the airport network[39], and the Petser online social network[49]. The plots show the average temporal distance or expected propagation time $D_{ij}$ for each pair of nodes $i,j$, and therefore encompass information about the expected spreading dynamics. In particular, we have ordered the nodes by their *characteristic spreading timescale*, which we define as the expected time after which a given node infects $\overline{N}$ other nodes, $\tau_i$, as

$$\tau_i = t|_{\sum_{j \neq i} \Theta(t - D_{ij}) = \overline{N}},$$
(8)

where $\Theta(\cdot)$ denotes the Heaviside step function, which is one when $t \geq D_{ij}$ and zero otherwise. A small value of $\tau_i$ means that node $i$ is of high importance or influence. This ordering of nodes displays the smooth gradient of the average temporal distance between pairs of nodes on all three empirical networks.

**Source detection.** For a given snapshot $r^*$ of a spreading process at time $t_0$ on a network, how does one detect the source node that generated the snapshot? Source detection[18,19,36–39] is an important theoretical and practical problem in network science. It is at least as hard as any problem in NP-complete class[37,51], because it belongs to #P-complete class[52]. We approach the problem by probabilistic kernel estimations[36,53] of the likelihood that source node generated the observed snapshot. Our framework can be used to perform this task, as the ensemble of weighted networks encompasses information about different realizations of the spreading process from all possible sources.

For the observed realization $r^*$, our source probability estimates for every node $v_i$ are constructed from ensemble averages $s_i \propto \sum_j^m f(\mathcal{G}_j)$. From weighted network $\mathcal{G}_j$, we extract realization based on temporal distances and calculate the similarity $\varphi_j \in [0,1]$ to the observed realization $r^*$. Similarity $\varphi_j$ is the normalized number of equal corresponding states between two realizations. Finally, we use the Gaussian kernel function

$$f(\mathcal{G}_j) = exp(-(\varphi_j - 1)^2 / a^2)$$
(9)

for probability estimations[36,53] (see Supplementary information, Sec. 4 for more details). We compare our source probability estimations with direct Monte Carlo approach[36] and topological distance approach[54]. The direct Monte Carlo approach generates large number of simulations $n = 10^6 - 10^8$, from each potential source node $v_i$, and source probability estimation is $s_i \propto n_i$, where $n_i$ is the number of simulations that match the observed realization $r^*$. The topological distance approach calculates the average topological distance $\langle d_i \rangle$ from each potential source $v_i$ to other infected nodes in the observed realization $r^*$ and source estimation is $s_i \propto \langle d_i \rangle^{-1}$. The result is shown in Fig. 6, where source ranked probability estimates for different methods are shown. Indeed, the method based on the ensemble of weighted networks ($n = 10^5$ instances) performs better than a topological distance estimation[54] and correctly assigns a high probability to the source candidate identified by the direct Monte Carlo method[36].
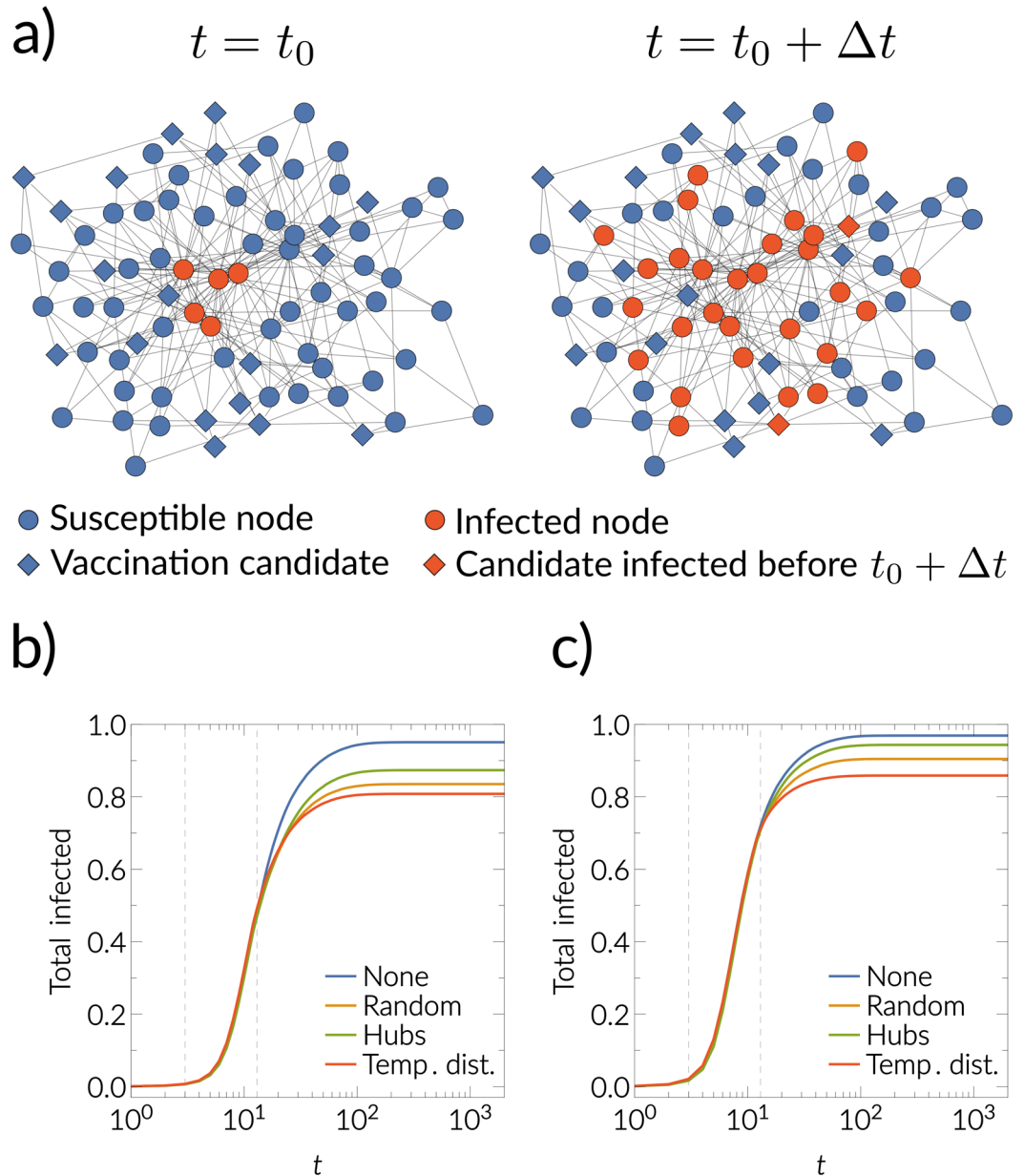
**Figure 7.** Time critical vaccination. (**a**) Shows an illustration of the time critical vaccination process. Vaccination candidates are highlighted by the diamond symbols. Red note denote infected individuals, and blue nodes are susceptible. In (**b**) and (**c**), we show the total number of infected nodes up to time $t$ for the different vaccination strategies described in the text ($m = 0.2$ of the total number of nodes). The first dashed vertical line denotes $T_0 = 3$, and the second one $T_0 + \tau = 13$. In (**b**) for the discrete time SIR $\beta = 0.03, \gamma = 0.01$ and in (**c**) for the discrete time SIR $\beta = 0.05, \gamma = 0.01$ with source node 10 in the Petster network.

**Time critical vaccination.** Now, let us consider the following vaccination scenario. We observe at time $t_0$ the outbreak of a disease and we have a certain amount, $m$, of vaccines, see Fig. 7(a). We can vaccinate $m$ individuals instantaneously, but the vaccines will only be effective after some time $\Delta t$ (a vaccinated individual can get infected before $t_0 + \Delta t$, and in this case this vaccine portion is wasted, see Fig. 7(b). Who shall we vaccinate? Our framework can be used to improve the vaccination strategy. The idea is that, at time $t_0$, we vaccinate only individuals who—with high probability—will not get infected before $t_0 + \Delta t$. Therefore, we estimate the probability that a node $i$ will not be infected before $t_0 + \Delta t$ as

$$\widetilde{p}_i = 1/n \sum_{k=1}^{n} \Theta(d_{\mathcal{G}_k}(v_\theta, v_i) - t_0 - \Delta t), \tag{10}$$

where $\theta$ denotes the source node and $\Theta(\cdot)$ is one when $d_{\mathcal{G}_k}(v_\theta, v_i) \geq t_0 + \Delta t$ and zero otherwise. We vaccinate nodes proportional to $\widetilde{p}$. In Fig. 7(b,c) we show using empirical Petster network that this procedure performs

better than randomly choosing from all susceptible nodes at $t_0$. Interestingly, vaccinating proportional to degree (hubs strategy) performs even worse than the fully random strategy, since the hubs usually are infected earlier, which increases the chance to waste vaccine portions.

## Discussion

We have mapped spreading dynamics on networks to an ensemble of weighted networks. This mapping allows us to define a temporal distance as shortest paths between nodes on an ensemble of weighted networks. We concentrate on the stochastic formulation of the generalized continuous and discrete time Susceptible Infected Recovered (SIR) spreading dynamics without memory (exponential inter-event distribution) and with memory (arbitrary inter-event distributions).

The problem of shortest path with the stochastic independent and identically distributed weights (i.i.d.) has been extensively researched in theoretical computer science[55–58], probability theory[59,60] and statistical physics community[31–35]. However, the general SIR dynamics is exactly represented with networks where the edge weights are conditionally independent random variables and only in special case of spreading they become independent variables e.g. SI dynamics ($\gamma = 0$) or SIR dynamics with fixed recovery time ($\psi(t) = \delta(t - T)$). Thus, our mapping to the ensemble of weighted networks, provides additional dynamical interpretation of the shortest paths problem in disordered networks. Previously, only the Independent Cascade spreading models (edge weights are i.i.d.) were connected to the problem of shortest paths[61] on networks, but with no connections to disordered networks or bond percolation literature.

In contrast to message passing, our framework is applicable for arbitrary network structures without neglecting loops in the network structure. Furthermore, our method takes into account dynamical correlations between node states overcoming the assumptions of mean-field like approximations[12,14]. In contrast to the well known historical Gillespie or kinetic Monte Carlo methods[23–28,36], we do not have to specify initial conditions upfront and we can sample new realizations from the previous ones by making local random perturbations on the weighted networks. We have shown that for some non-Markovian processes, the average propagation time can scale as a polynomial with the system size, even if the underlying network is small world. In stark contrast to the Markovian case, this means that for large enough systems the disease does not spread globally in a realistic timescale. We have shown that our framework allows for direct applications like ranking of nodes according to their characteristic spreading timescale, source detection and time-critical vaccination. Combining our mapping framework with existing vaccination strategies is likely to further improve the results, especially when time is a crucial factor.

Note that previous studies[39,62–64] for estimating the spreading arrival times were concentrated on a global disease spread with different approximations and theoretical understanding of effective distances[65] for metapopulation models. However, the exact mathematical equivalences for mapping generalized spreading dynamics at a microscopic level were still lacking. Furthermore, in a limit of infinite process time, we establish the connection with bond percolation theory. Previously the outcome in infinite time of the SIR process have been successfully mapped to the percolation[20,21,29,30] but not for the finite time. Contrary to previous state-of-the-art source inference studies[18,19,36], estimations with our method require same computational effort to sample realizations at an arbitrary point in time $t$ not only for discrete time but also for continuous time dynamics.

## References

1. Guille, A., Hacid, H., Favre, C. & Zighed, D. A. Information diffusion in online social networks. *ACM SIGMOD Record* **42**, 17 (2013).
2. Lerman, K. & Ghosh, R. Information contagion: an empirical study of the spread of news on digg and twitter social networks. In *in Proc. 4th Int. Conf. on Weblogs and Social Media (ICWSM)*, 2010.
3. Anderson, R. M. & May, R. M. *Infectious Diseases in Humans* (Oxford University Press, 1992).
4. Vespignani, A. Modelling dynamical processes in complex socio-technical systems. *Nature Physics* **8**, 32–39 (2011).
5. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
6. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D.-U. Complex networks: Structure and dynamics. *Physics Reports* **424**, 175–308 (2006).
7. Barrat, A., Barthélemy, M. & Vespignani, A. *Dynamical Processes on Complex Networks* (Cambridge University Press, 2012).
8. Gillespie, D. T. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications* **188**, 404–425 (1992).
9. Van Mieghem, P. & Cator, E. Epidemics in networks with nodal self-infection and the epidemic threshold. *Phys. Rev. E* **86**, 016116 (2012).
10. Van Mieghem, P., Omic, J. & Kooij, R. Virus spread in networks. *IEEE/ACM Transactions on Networking* **17**, 1–14 (2009).
11. Gleeson, J. P. Binary-state dynamics on complex networks: Pair approximation and beyond. *Phys. Rev. X* **3**, 021004 (2013).
12. Castellano, C. & Pastor-Satorras, R. Thresholds for epidemic spreading in networks. *Phys. Rev. Lett.* **105**, 218701 (2010).
13. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203 (2001).
14. Sharkey, K. J. Deterministic epidemic models on contact networks: Correlations and unbiological terms. *Theoretical Population Biology* **79**, 115–129 (2011).
15. Sharkey, K. J., Kiss, I. Z., Wilkinson, R. R. & Simon, P. L. Exact equations for sir epidemics on tree graphs. *Bulletin of Mathematical Biology* **77**, 614–645 (2015).
16. Kiss, I. Z., Morris, C. G., Sélley, F., Simon, P. L. & Wilkinson, R. R. Exact deterministic representation of markovian sir epidemics on networks with and without loops. *Journal of Mathematical Biology* **70**, 437–464 (2015).
17. Karrer, B. & Newman, M. E. J. Message passing approach for general epidemic models. *Phys. Rev. E* **82**, 016101 (2010).
18. Lokhov, A. Y., Mézard, M., Ohta, H. & Zdeborová, L. Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Phys. Rev. E* **90**, 012801 (2014).
19. Altarelli, F., Braunstein, A., Dall'Asta, L., Lage-Castellanos, A. & Zecchina, R. Bayesian inference of epidemics on networks via belief propagation. *Phys. Rev. Lett.* **112**, 118701 (2014).
20. Mollison, D. Spatial contact models for ecological and epidemic spread. *J R Stat Soc Series B* **39**, 283–326 (1977).
21. Grassberger, P. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences* **63**, 157–172 (1983).
22. Meyers, L., Newman, M. & Pourbohloul, B. Predicting epidemics on directed contact networks. *Journal of Theoretical Biology* **240**, 400–418 (2006).

23. Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* **22**, 403–434 (1976).
24. Fichthorn, K. A. & Weinberg, W. H. Theoretical foundations of dynamical monte carlo simulations. *The Journal of Chemical Physics* **95**, 1090 (1991).
25. Bortz, A., Kalos, M. & Lebowitz, J. A new algorithm for monte carlo simulation of ising spin systems. *Journal of Computational Physics* **17**, 10–18 (1975).
26. Boguñá, M., Lafuerza, L. F., Toral, R. & Serrano, M. A. Simulating non-markovian stochastic processes. *Phys. Rev. E* **90**, 042108 (2014).
27. Vestergaard, C. L. & Génois, M. Temporal gillespie algorithm: fast simulation of contagion processes on time-varying networks. *PLoS Comput Biol* **11**, 1–28 (2015).
28. Antulov-Fantulin, N., Lancic, A., Stefancic, H. & Sikic, M. FastSIR algorithm: A fast algorithm for the simulation of the epidemic spread in large networks by using the susceptible–infected–recovered compartment model. *Information Sciences* **239**, 226–240 (2013).
29. Newman, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016128 (2002).
30. Kenah, E. & Robins, J. M. Second look at the spread of epidemics on networks. *Phys. Rev. E* **76**, 036113 (2007).
31. Chen, Y., López, E., Havlin, S. & Stanley, H. E. Universal behavior of optimal paths in weighted networks with general disorder. *Phys. Rev. Lett.* **96**, 068702 (2006).
32. Braunstein, L. A., Buldyrev, S. V., Cohen, R., Havlin, S. & Stanley, H. E. Optimal paths in disordered complex networks. *Phys. Rev. Lett.* **91**, 168701 (2003).
33. Sreenivasan, S. *et al.* Effect of disorder strength on optimal paths in complex networks. *Phys. Rev. E* **70**, 046133 (2004).
34. Van Mieghem, P. & van Langen, S. Influence of the link weight structure on the shortest path. *Phys. Rev. E* **71**, 056113 (2005).
35. Van Mieghem, P. & Magdalena, S. M. Phase transition in the link weight structure of networks. *Phys. Rev. E* **72**, 056138 (2005).
36. Antulov-Fantulin, N., Lančić, A., Šmuc, T., Štefančić, H. & Šikić, M. Identification of patient zero in static and temporal networks: Robustness and limitations. *Phys. Rev. Lett.* **114**, 248701 (2015).
37. Shah, D. & Zaman, T. Detecting sources of computer viruses in networks: theory and experiment. In *Proceedings of the ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, SIGMETRICS '10, 203–214 (ACM, New York, NY, USA, 2010).
38. Pinto, P. C., Thiran, P. & Vetterli, M. Locating the Source of Diffusion in Large-Scale Networks. *Phys. Rev. Lett.* **109**, 068702+ (2012).
39. Brockmann, D. & Helbing, D. The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science* **342**, 1337–1342 (2013).
40. Jalili, M. & Perc, M. Information cascades in complex networks. *Journal of Complex Networks* **5**, 665–693 (2017).
41. Wang, Z. *et al.* Statistical physics of vaccination. *Physics Reports* **664**, 1–113 (2016).
42. Helbing, D. *et al.* Saving human lives: What complexity science and information systems can contribute. *Journal of Statistical Physics* **158**, 735–781 (2015).
43. Devroye, L. Sample-based non-uniform random variate generation. In *Proceedings of the 18th Conference on Winter Simulation*, WSC '86, 260–265 (ACM, New York, NY, USA, 1986).
44. Dijkstra, E. W. A note on two problems in connexion with graphs. *Numerische Mathematik* **1**, 269–271 (1959).
45. Demetrescu, C. & Italiano, G. F. Algorithmic techniques for maintaining shortest routes in dynamic networks. *Electronic Notes in Theoretical Computer Science* **171**, 3–15 Proceedings of the Second Workshop on Cryptography for Ad-hoc Networks (WCAN 2006) (2007).
46. Demetrescu, C. & Italiano, G. F. A new approach to dynamic all pairs shortest paths. *J. ACM* **51**, 968–992 (2004).
47. Starnini, M., Gleeson, J. P. & Boguna, M. Equivalence between non-markovian and markovian dynamics in epidemic spreading processes. *Phys. Rev. Lett.* **118**, 128301 (2017).
48. Guimerà, R., Danon, L., Díaz-Guilera, A., Giralt, F. & Arenas, A. Self-similar community structure in a network of human interactions. *Phys. Rev. E* **68**, 065103 (2003).
49. Rossi, R. A. & Ahmed, N. K. The network data repository with interactive graph analytics and visualization. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence* (2015).
50. Woolley-Meza, O. *et al.* Complexity in human transportation networks: a comparative analysis of worldwide air transportation and global cargo-ship movements. *The Eur. Phys. Journal B - Condensed Matter and Complex Systems* **84**, 589–600 (2011).
51. Zhai, X., Wu, W. & Xu, W. Cascade source inference in networks: a markov chain monte carlo approach. *Computational Social Networks* **2**, 17 (2015).
52. Provan, J. S. & Ball, M. O. The complexity of counting cuts and of computing the probability that a graph is connected. *SIAM Journal on Computing* **12**, 777–788 (1983).
53. Marron, J. S. & Nolan, D. Canonical kernels for density estimation. *Statistics & Probability Letters* **7**, 195–199 (1988).
54. Comin, C. H. & da F Costa, L. Identifying the starting point of a spreading process in complex networks. *Phys. Rev. E* **84**, 056105 (2011).
55. Valiant, L. G. The complexity of enumeration and reliability problems. *SIAM Journal on Computing* **8**, 410–421 (1979).
56. Kulkarni, V. G. Shortest paths in networks with exponentially distributed arc lengths. *Networks* **16**, 255–274 (1986).
57. Corea, G. A. & Kulkarni, V. G. Shortest paths in stochastic networks with ARC lengths having discrete distributions. *Networks* **23**, 175–183 (1993).
58. Peer, S. & Sharma, D. K. Finding the shortest path in stochastic networks. *Computers & Mathematics with Applications* **53**, 729–740 (2007).
59. Hofstad, R., van der, Hooghiemstra, G. & Van Mieghem, P. Size and weight of shortest path trees with exponential link weights. *Combinatorics, Probability and Computing* **15**, 903 (2006).
60. Bhamidi, S., Goodman, J., Hofstad, Rvander & Komjáthy, J. Degree distribution of shortest path trees and bias of network sampling algorithms. *The Annals of Applied Probability* **25**, 1780–1826 (2015).
61. Gomez-Rodriguez, M., Song, L., Du, N., Zha, H. & Schölkopf, B. Influence estimation and maximization in continuous-time diffusion networks. *ACM Transactions on Information Systems* **34**, 1–33 (2016).
62. Gautreau, A., Barrat, A. & Barthélemy, M. Global disease spread: statistics and estimation of arrival times. *Journal of Theoretical Biology* **251**, 509–522 (2008).
63. Colizza, V. & Vespignani, A. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *Journal of Theoretical Biology* **251**, 450–467 (2008).
64. Lawyer, G. Measuring the potential of individual airports for pandemic spread over the world airline network. *BMC Infectious Diseases* **16**, 70 (2016).
65. Iannelli, F., Koher, A., Brockmann, D., Hövel, P. & Sokolov, I. M. Effective distances for epidemics spreading on complex networks. *Phys. Rev. E* **95**, 012313 (2017).

## Acknowledgements

## Author Contributions

All authors contributed to the writing and editing of the manuscript, proofreading and analyses. D.T. contributed to defining, formalizing and generalizing the mapping, relation to bond percolation and applications, K.K.K. contributed to defining temporal maps, time critical vaccination and connection to disordered networks and N.A.F. contributed to probabilistic modelling, source inference, algorithmic contributions and lead the research.

## Additional Information

**Supplementary information** accompanies this paper at https://doi.org/10.1038/s41598-018-24648-w.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.