Check for
updates

Regular Article

Path-integral formulation of spreading processes in complex networks

Flavio Iannelli[1,a] and Igor M. Sokolov[1]

[1] Institute for Physics, Humboldt-University of Berlin, Newtonstrasse 15, 12489 Berlin, Germany

Abstract We introduce a path-integral formulation of network-based measures that generalize the concept of geodesic distance and that provides fundamental insights into the dynamics of disease transmission as well as an efficient numerical estimation of the infection arrival time.

1 Introduction

The forecast and control of emergent diseases has become particularly important in recent years because of the increasingly growing structure and velocity of transportation means. A series of papers have been devoted to the problem of infection arrival times in complex networks [2,3,6,7]. Most of them are based on the assumption that a single dominant path, associated with maximal traffic probability, is sufficient to estimate accurately the arrival time of a diffusive process. While this is partially true in some specific cases, when a single path between each pair of nodes is available, in the general scenario this assumption can give poor estimates for the arrival time.

Effective distances (ED) in the dominant-path approach can be defined, for both directed and undirected networks, as the geodesic graph distance of a weighted graph with edge weights given by the first moment of a distribution known from extreme events statistics [8], which depends only on the network topology and on the transmission and recovery rates. This approach has the disadvantage that it can significantly overestimate the infection arrival time obtained numerically [5,6]. In addition, in situations where multiple equiprobable paths exist between node pairs, as in regular lattices, this approach breaks down. A more realistic scenario takes into account all possible propagation routes [7] yielding a multiple-path ED, which is supposedly the best possible estimate of infection arrival times. Unfortunately, the computation becomes infeasible as the number of paths between two nodes grows exponentially with the size of the network. The lack of a practical computational approach leads back to considering only the dominant path.

Here, we introduce a path-integral formulation that generalizes the multiple-path ED and that can be used

as a computationally feasible alternative with a clear interpretation, paving the way for future studies on spreading processes in complex networks.

The paper is organized as follows. In Sect. 2.1, we define the dominant-path ED by considering node-independent exit rates. In Sect. 2.2, we present the generalization to multiple paths of arbitrary length connecting two nodes. In Sect. 2.3, we follow the random-walk approach defining the path-integral formulation of the problem. Then using a metapopulation model, we compare the path-integral formulation of ED with the infection arrival times. In Sect. 3, we summarize our results and present our conclusions.

2 Network effective distances

2.1 Dominant path

The fundamental metric in networks is the geodesic distance, i.e. the shortest-path length over all paths $\{\Gamma_{ij}\}$ connecting node i to node j. For weighted networks, where the weights W_{ij} are positive numbers identifying the carrying capacity of a certain route, each edge $(k,l) \in \Gamma_{ij}$ contributes to the total length with its reciprocal weight. This generalizes the standard definition for unweighted networks as

$$D_{ij} = \min_{\{\Gamma_{ij}\}} \sum_{(k,l) \in \Gamma_{ij}} W_{kl}^{-1}. \tag{1}$$

The reciprocal of the weights is used consistently with the fact that a higher flux of passengers along an edge reduces the distance between the respective nodes.

Starting from the heuristic definition equation (1), it is possible to extend the notion of distance by replacing the weights with an effective function of the weights $f(W_{kl})$ that quantitatively reproduces the distance as measured by the arrival time of spreading processes

a e-mail: flavio.iannelli@business.uzh (corresponding author)

unfolding on a given network. In [3], the authors proposed the ansatz

$$D_{ij}^{\text{eff}} = \min_{\{\Gamma_{ij}\}} \sum_{(k,l)\in\Gamma_{ij}} (1 - \ln P_{kl}), \qquad (2)$$

where $P_{kl} = W_{kl}/\sum_m W_{km}$ is the transition probability to navigate the graph via random walks. The choice for the logarithm is motivated by the authors simply by requiring the additivity of the distance equation (2), consistently with multiplying the corresponding probabilities. Although this ED is able to reproduce to a good approximation the infection arrival time in the GMN, its interpretation is quite obscure and the particular choice of the logarithm function is not derived from a microscopic description of the process. Furthermore, the choice for the constant term equal to unity in its definition is arbitrary and it is not clear a priori why such expression would correctly quantify the arrival times of reaction–diffusion processes. Most importantly, the definition equation (2) suffers from the bias of considering a single (probability-dominant) path, neglecting all other routes for information transfer.

An alternative and refined approach to derive analytically the ED using a detailed kinetic description of spreading processes was outlined in [7]. Interestingly, Eq. (2) can be obtained as a special case of a more general quantity. To show this, one needs to extend and generalize the Markovian description presented in [7] and define the ED using a dominant-path approach, by requiring the maximization of the travel probability between adjacent subpopulations, which can be expressed in terms of the matrix P_{ij}. The main difference from the existing result is that we assume that the exit rate of each node $q_i = \sum_{l\neq i} Q_{il}$ is independent on the node location, i.e. $q_i = \alpha$ for all nodes. After some rearrangements, one obtains an expression with the same form of Eq. (2), but with an additional dependence on the spreading parameters [9].

This yields the dominant-path ED

$$D_{ij}^{\text{DP}}(\lambda) = \min_{\{\Gamma_{ij}\}} \sum_{(k,l)\in\Gamma_{ij}} (\lambda - \ln P_{kl}). \qquad (3)$$

Here we have defined

$$\lambda = \ln \frac{\beta - \mu}{\alpha} - \gamma_e, \qquad (4)$$

where β and μ are the infection and recovery rates while $\gamma_e \approx 0.577$ is the Euler–Mascheroni constant. From Eq. (3), we can recover the ansatz for the ED equation (2) simply by setting $\lambda = 1$. However, in the dominant-path ED equation (3), the definition of λ as a function of the epidemic and mobility parameters gives the optimized edge weight that should contribute into the minimization condition over all paths connecting source and target. On the computational side, one can obtain the full matrix D_{ij}^{DP} using the Dijkstra algorithm

in a time $\mathcal{O}(NE + N^2 \log N)$, where E and N are the graph size and order, respectively.

The most important limitation of Eq. (3) is that only the path that minimizes the topological length and at same time maximizes the associated probability is considered. It turns out that, because of this limitation, the infection arrival time $D_{ij}^{\text{DP}}(\lambda)/v$, where $v \approx \beta - \mu$ is the linearized speed of the infection [3], is overestimated with respect to the numerical arrival times obtained from direct simulations [5,6]

2.2 Multiple paths

The correct approach is to consider the multiplicity of transmission routes. The framework to include all directed paths of transmission was developed in [7]. For simplicity, we start by considering only two distinct paths Γ and Γ' connecting the same pair of nodes. The two-path ED D_{ij}^{2P} that generalizes the dominant-path ED equation (3) satisfies the equation

$$e^{-D_{ij}^{\text{2P}}} = e^{-D_{ij}^{\Gamma}} + e^{-D_{ij}^{\Gamma'}}, \qquad (5)$$

where

$$D_{ij}^{\Gamma}(\lambda) = \sum_{(k,l)\in\Gamma_{ij}} (\lambda - \ln P_{kl}) = -\ln\left(\prod_{(k,l)\in\Gamma_{ij}} e^{-\lambda}P_{kl}\right) \qquad (6)$$

is the ED, which is the mean of a Gumbel distribution, associated with a path Γ_{ij} of arbitrary length connecting node i to node j. Equation (5) can be easily generalized to an arbitrary number of paths connecting the same pair of nodes, as

$$\exp\left(-D_{ij}^{\text{MP}}(\lambda)\right) = \sum_{\{\Gamma_{ij}\}} \exp\left(-D_{ij}^{\Gamma}(\lambda)\right). \qquad (7)$$

The previous equation defines the multiple-path ED

$$D_{ij}^{\text{MP}}(\lambda) = -\ln\left(\sum_{\{\Gamma_{ij}\}} e^{-\lambda n(\Gamma)}P(\Gamma)\right), \qquad (8)$$

where the total probability associated to the path Γ_{ij} of length $n(\Gamma_{ij}) = |\Gamma_{ij}|$ is

$$P(\Gamma_{ij}) = \prod_{(k,l)\in\Gamma_{ij}} P_{kl}. \qquad (9)$$

An analogous expression can be obtained by grouping all probabilities associated to paths of same length into the quantity $F_{ij}(n) = \sum_{|\Gamma|=n} P(\Gamma_{ij})$. Then we can replace the sum over all paths connecting i to j in Eq. (8) with a sum over the allowed path lengths

$n = |\Gamma| \in [1, n_{\max}]$, to get

$$D_{ij}^{\mathrm{MP}}(\lambda) = -\ln\left(\sum_{n=1}^{n_{\max}} e^{-\lambda n} F_{ij}(n)\right), \qquad (10)$$

where $n_{\max}$ is the maximum path length in the network. If instead of considering all paths in Eq. (8) we select the single path Γ_{ij}^* of length $n(\Gamma_{ij}^*) = |\Gamma_{ij}^*|$ that is associated to the dominant contribution, i.e. the path that maximizes its associated probability and minimizes the topological path length, we recover the dominant-path ED equation (3), i.e.

$$D_{ij}^{\mathrm{DP}}(\lambda) = D_{ij}^{\Gamma^*}(\lambda) = -\ln\left(e^{-\lambda n(\Gamma_{ij}^*)} P(\Gamma_{ij}^*)\right). \qquad (11)$$

The multiple-path ED provides an accurate estimate of the infection arrival time, as it counts the most probable route as well as all possible alternative-directed transmission routes. However, a major drawback of this approach is that it is computationally not tractable. In fact, the total number of paths between i and j is bounded but exponentially growing with the number of nodes N, thus the measure D_{ij}^{MP} becomes useless for large networks [13]. A trade-off between performance and accuracy can be achieved by restricting the path search algorithm to a maximum path length or more elegantly, as we show next, by relaxing the assumption of direct propagation and define a path-integral formulation of the problem.

2.3 Random walks

Both measures introduced in the previous sections D_{ij}^{DP} and D_{ij}^{MP} rely on the fact that the epidemic will spread along paths, i.e. routes that do not ever cross. Here we follow a different approach and introduce an ED that includes all possible random walks from source to target. Relaxing the assumption of directed spread is equivalent to effectively erasing the memory from the system at each time step. This is achieved by including in Eq. (8) all walks $\{\Xi\}$ that, contrary to the paths $\{\Gamma\}$, allow also walking on already visited nodes. We define the random-walk effective distance (RWED) by generalizing Eq. (8) as

$$D_{ij}^{\mathrm{RW}}(\lambda) = -\ln\left(\sum_{\{\Xi_{ij}\}} e^{-\lambda n(\Xi)} P(\Xi)\right), \qquad (12)$$

where $P(\Xi_{ij}) = \prod_{(k,l) \in \Xi_{ij}} P_{kl}$ is the total probability associated to the walk Ξ_{ij} of length $n(\Xi_{ij}) = |\Xi_{ij}|$. We note that since $\{\Xi\}$ is a bigger set than $\{\Gamma\}$, the following inequalities hold:

$$D_{ij}^{\mathrm{RW}} \leq D_{ij}^{\mathrm{MP}} \leq D_{ij}^{\mathrm{DP}}. \qquad (13)$$

As we did in the previous section for paths with probability $P(\Gamma)$, we can group the probabilities associated to walks of the same length into the quantity $H_{ij}(n) = \sum_{|\Xi|=n} H_{ij}(\Xi_{ij})$. The latter is precisely the hitting time probability defined recursively for $i \neq j$ as $H_{ij}(n) = \sum_{k \neq j} P_{ik} H_{kj}(n)$. Contrary to the multiple-path scenario, walks are unbounded and so becomes the maximum length $n_{\max}$ in Eq. (10). Substituting, we rewrite the RWED as

$$D_{ij}^{\mathrm{RW}}(\lambda) = -\ln\left(\sum_{n=1}^{\infty} e^{-\lambda n} H_{ij}(n)\right). \qquad (14)$$

Remarkably, there is a immediate interpretation of the RWED in terms of the hitting-times generating functions. Indeed by definition

$$D_{ij}^{\mathrm{RW}}(\lambda) = -\ln\left\langle e^{-\lambda n_{ij}}\right\rangle, \qquad (15)$$

where n_{ij} is the random-walk hitting time to node j [10] and the average is taken over all random walks that start in i and terminate as soon as they hit node j. It is easy to see that

$$D_{ij}^{\mathrm{RW}}(\lambda) = \sum_{k=0}^{\infty} (-1)^{k+1} \frac{\lambda^k}{k!} \left\langle n_{ij}^k\right\rangle_c = -\Psi_{ij}(-\lambda), \qquad (16)$$

where $\left\langle n^k\right\rangle_c$ are the hitting time cumulants and $\Psi_{ij}(\lambda)$ is the cumulant generating function of the random-walk hitting time. Hence, one obtains the cumulants of the random-walk hitting time by differentiating Eq. (14) with respect to λ. In particular, the mean-first-passage time (MFPT) from i to j is obtained as $M_{ij} = \left\langle n_{ij}\right\rangle_c = \partial_\lambda D_{ij}^{\mathrm{RW}}|_{\lambda=0}$.

To compute $D_{ij}^{\mathrm{RW}}(\lambda)$, we can write $H_{ij}(n)$ in terms of powers of the transition probability sub-matrix $\mathbf{P}^{(j)}$ obtained by removing the jth row and column, as for the computation of the MFPT. Assuming a positive λ, the expansion in a geometric series converges[1] and we obtain

$$D_{ij}^{\mathrm{RW}}(\lambda) = -\ln\left(\sum_{k \neq j}\left(\mathbf{I}^{(j)} - e^{-\lambda}\mathbf{P}^{(j)}\right)^{-1}_{ik} e^{-\lambda} p_k^{(j)}\right). \qquad (17)$$

Here, $\mathbf{I}^{(j)}$ is the $(N-1) \times (N-1)$ sub-identity matrix and $\mathbf{p}^{(j)}$ is the jth column of $\mathbf{P}$ with jth component removed that takes into account the last step needed to reach the target j.

---

[1] To have a converging expression for Eq. (14), we must require $\lambda > 0$ and then the determinant of the matrix $e^{-\lambda}\mathbf{P}^{(j)}$ is always smaller than unity. By recalling the definition equation (2.1), the previous condition imposes an additional constraint on the model parameters, i.e. $\beta > \mu + \alpha e^{-\gamma_e} \approx \mu + 2\alpha$.

Next we show how we can disentangle the dominant contribution defined by D_{ij}^{DP} in the RWED, using a path-integral formulation [14] of ED. Indeed, we can rewrite the RWED equation (12) as

$$D_{ij}^{\mathrm{RW}} = -\ln\left(\sum_{\{\Xi_{ij}\}} e^{-\mathcal{A}(\Xi,\lambda)}\right), \qquad (18)$$

where $\mathcal{A}(\Xi,\lambda)$ is interpreted as the Euclidean action defined as

$$\mathcal{A}(\Xi,\lambda) = \lambda n(\Xi) - \ln P(\Xi). \qquad (19)$$

As previously, $n(\Xi)$ is the length of the walk Ξ and $P(\Xi)$ is the associated probability. In this picture, the RWED is the the free energy functional (changed in sign) of an Euclidean field theory [11,14]. The dominant-path ED is instead described by Eq. (11). In contrast to Eq. (18), the dominant-path ED neglects contributions of paths other than the one that maximizes the associated probability. Then we can think of D_{ij}^{DP} as an *effective action*[2]

$$D_{ij}^{\mathrm{DP}} = \mathcal{A}^{\mathrm{eff}}(\Gamma_{ij}^*,\lambda) = \lambda n(\Gamma_{ij}^*) - \ln P(\Gamma_{ij}^*), \quad (20)$$

where Γ_{ij}^* is the (probability) dominant path of length $n(\Gamma_{ij}^*)$. Expanding the action around the dominant contribution as

$$\mathcal{A}(\Xi,\lambda) \approx \mathcal{A}^{\mathrm{eff}}(\Gamma^*,\lambda) + \left.\frac{\delta\mathcal{A}}{\delta\Xi}\right|_{\Xi=\Gamma^*}\delta\Xi + \cdots \quad (21)$$

and plugging this into Eq. (18) we find

$$D_{ij}^{\mathrm{RW}} = \mathcal{A}^{\mathrm{eff}}(\Gamma_{ij}^*,\lambda) - \ln\left(\sum_{\{\Xi_{ij}\}} e^{\left.\frac{\delta\mathcal{A}}{\delta\Xi}\right|_{\Xi=\Gamma^*}\delta\Xi + \dots}\right). \qquad (22)$$

In the previous equation $\delta\Xi$ quantifies the deviation of the walk Ξ with respect to the dominant path Γ^*. By comparing with the ansatz D_{ij}^{eff} defined by Eq. (2), we can finally identify

$$D_{ij}^{\mathrm{eff}} = \mathcal{A}^{\mathrm{eff}}(\Gamma_{ij}^*,1), \qquad (23)$$

where the effective action is defined by Eq. (20). The additional contributions in the RWED with respect to the effective action, given by the logarithmic term on the right-hand side of Eq. (22), account for the multiplicity of transmission routes. The latter quantifies the discrepancy between the random-walk and dominant-path approach. We expect this contributions

---

² Note that our definition is not equivalent to the effective action of quantum field theory used to analyze the renormalizability of the theory [12].

to be negligible for network topologies that are locally tree-like, where a single path connects any pair of nodes. For many real-world networks, this is the case since the degree distribution behaves as a power-law and the navigation is dominated by the presence of large hubs [4]. However, we expect the dominant-path approach to completely break down when the multiplicity of paths becomes relevant as in the case of random (Poissonian) networks and regular lattices where a large number of paths are equally probable. In this case, the difference $|D_{ij}^{\mathrm{RW}} - D_{ij}^{\mathrm{DP}}|$ will be non-negligible and the arrival time estimates will substantially decrease when considering the random-walk approach.

As the substrate for epidemic spreading, we consider the global mobility network (GMN), where nodes are different airports and weighted edges represents the number of seats on scheduled commercial flights between them. In Fig. 1, we show the GMN by drawing an arc for each link connecting two airports, color coded according to the flux of passengers on that connection. Assuming that the number of seats on scheduled commercial flights is on average proportional to the number of passengers traveling, the data can be represented as a weighted network with adjacency matrix W_{ij} giving the total traffic per day between airport i and airport j.

To each airport j (a node in the metapopulation) is associated a subpopulation of size N_j so that $W_{ij} = Q_{ij}N_i$, where Q_{ij} is the transition rate from i to j. Although the network is directed, the degree of asymmetry in the weighted adjacency matrix W_{ij} is very low. This particular feature is well confirmed also in similar empirical datasets of air-traffic at the global scale [1,3]. We estimate quantitatively the degree of asymmetry by looking at both the topological and weights asymmetry. The former is quantified by the average number of non-zero elements ϵ in the corresponding unweighted adjacency matrix $A_{ij} = \chi(W_{ij})$, where χ is the step function equal to one for positive arguments and vanishing otherwise. Instead the weight asymmetry ϵ_w is defined as the normalized net difference between travel fluxes in each route and the corresponding reversed travel. We find $\epsilon = 2 \cdot 10^{-3}$ and $\epsilon_w = 3.1 \cdot 10^{-9}$. Thus, we redefine the a symmetric adjacency matrix as $W_{ij} = (W_{ij} + W_{ji})/2$. Symmetrizing the weights also assures us that the subpopulation sizes are conserved quantities since $\dot{N}_i = \sum_k (W_{ki} - W_{ik})$. In principle, the matrix W_{ij} is provided by traffic data and N_i by census data such that the rates Q_{ij}, for $i \neq j$, can be computed inverting $W_{ij} = Q_{ij}N_i$. However, although it is straightforward to measure W_{ij}, assessing the effective population is more subtle. The number of individuals that effectively participate in the dispersal N_i is not necessarily the same as the population data provided by census. We also assume that the exit rate $q_i = \sum_{l\neq i} Q_{il}$ is independent of the node i and reformulate the diffusion contribution in terms of the transition matrix P_{ij} and a constant diffusion rate α. The latter is given by the

Fig. 1 The global mobility network (GMN) of air-traffic. Each edge corresponds to a scheduled commercial flight, with gradient scaling from dark to light-blue according to the available number of seats

ratio $\alpha = \mathcal{W}/\mathcal{N}$, between the total flux (per unit time) $\mathcal{W} = \sum_{ij} W_{ij}$ and the total population $\mathcal{N} = \sum_i N_i$. In terms of α, we can remove the dependence on the subpopulation size N_i from the reaction–diffusion equations.

For the numerical simulations, we use the SIR metapopulation model with transmission and recovery probability per time step given by the rates β and μ, respectively. We denote by S_j, I_j and R_j the number of individuals in subpopulation j who are susceptible, infected, and removed, respectively. The normalized quantities are $\rho_j^X = X_j/N_j$, where $X \in \{S, I, R\}$, and N_j is the subpopulation size. The full SIR metapopulation model is obtained by combining a diffusion term $\Omega(\{\rho^X\}) = \alpha \sum_k P_{ik} \left(\rho_k^X - \rho_i^X\right)$, with a global exit rate α, with the compartmental SIR reactions in a unified picture that reads

$$\dot{\rho}_i^S = \Omega(\{\rho^S\}) - \beta \rho_i^S \rho_i^I \tag{24}$$

$$\dot{\rho}_i^I = \Omega(\{\rho^I\}) + \beta \rho_i^S \rho_i^I - \mu \rho_i^I. \tag{25}$$

The key quantity we are interested in estimating is the infection arrival time, defined for each pair of node by the matrix

$$T_{ij} = \min_t \left\{ t \geq 0 \mid \rho_j^I(t) \geq 1/N_j \right\}_i, \tag{26}$$

where the subscript i implies that the process started in node i at time $t = 0$, and by definition $T_{ii} = 0$. We use T_{ij} as the benchmark of infection arrival times in real-world pandemic scenarios and compare it to the ED defined by Eq. (3) and Eq. (17). A comparison between RWED and infection arrival time is shown in Fig. 2. There we show four different time snapshots of a pandemic in the GMN with basic reproductive number $\mathcal{R}_0 = 1.5$ originated at São Paulo Guarulhos International Airport. The embedding of the GMN provided by the path-integral formulation equation (18) reveals simple circular wave centered at the seed of

the process, which are not present in the geographical space.

3 Discussion

In this paper, we have introduced a path-integral formulation of network effective distances by relaxing the assumption of simple-path propagation of spreading processes. These results provide additional insights into the relation between random-walk-based metrics and epidemic spreading in complex networks. The proposed RWED includes the previously defined dominant-path measure as a particular case. The almost perfect correlation found with the infection arrival time can be explained as follows. The contribution of looped trajectories for diffusion is dramatically reduced because of the decreasing exponential in the walk length in Eq. (12). The latter damps all contributions of very long walks, and in particular allows us to neglect the contributions of infinite loops. In scenarios where multiple parallel paths are important, for instance in Erdos–Renyi graphs or regular lattices, the assumption of a single dominant path breaks down and the measure proposed here can be used as an efficient alternative. The predictive power of the RWED can be used for containment strategies and estimation of arrival times for real global pandemics from the underlying networks topology. The method can in fact be generally applied to any weighted and directed network besides transportation networks, e.g. social networks. It remains intriguing to test if this is the case also for rumor spreading dynamics. For unweighted locally tree-like networks both the dominant-path ED and RWED yield maximum correlation with the simulated arrival time, as the dominant path tends to dominate. Our results show that infection arrival time in networked systems and heuristic definitions of ED can be obtained within a path-integral formulation of the
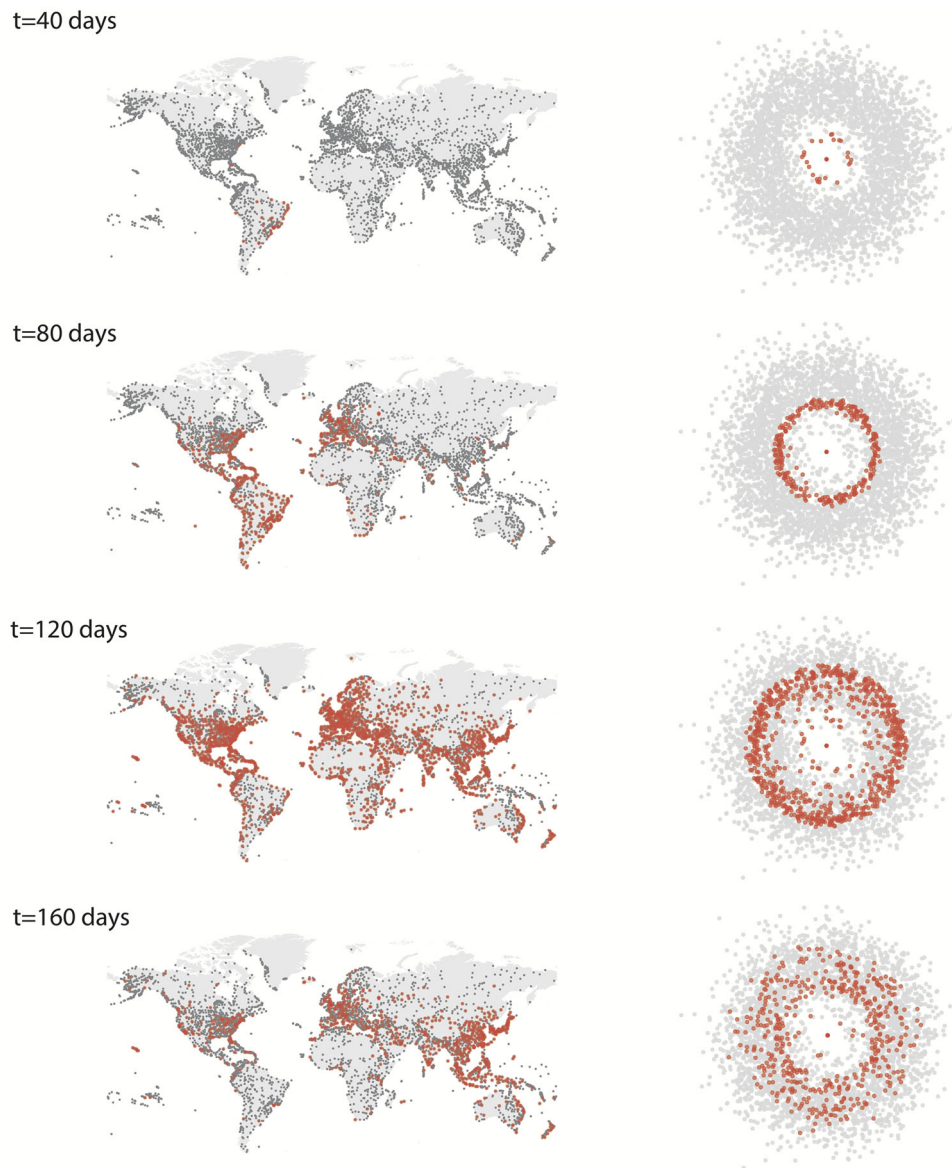
Fig. 2 Left: prevalence of a global pandemic with basic reproductive number $\mathcal{R}_0 = 1.5$ at four different observation times, as obtained from numerical integration of 25. The infection seed is São Paulo Guarulhos International Airport.

Right: corresponding plot in the hidden space of RWED, where the epidemic spreads as a highly correlated circular wave centered at the infection seed

problem and with the definition of suitable effective actions.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: This is a theoretical study and no experimental data.]

References

1. A. Barrat, M. Barthelemy, A. Vespignani. The architecture of complex weighted networks: measurements and models. In: *Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science* (World Scientific, 2007), pp. 67–92

2. L.A. Braunstein, S.V. Buldyrev, R. Cohen, S. Havlin, H.E. Stanley, Optimal paths in disordered complex networks. Phys. Rev. Lett. **91**, 168701 (2003)

3. D. Brockmann, D. Helbing, The hidden geometry of complex, network-driven contagion phenomena. Science **342**(6164), 1337–1342 (2013)

4. G. Caldarelli, *Scale-Free Networks* (Oxford University Press, Oxford, 2007)

5. P. Crepey, F.P. Alvarez, M. Barthélemy, Epidemic variability in complex networks. Phys. Rev. E **73**(4), 046131 (2006)

6. A. Gautreau, A. Barrat, M. Barthélemy, Arrival time statistics in global disease spread. J. Stat. Mech. Theory Exp. **2007**(09), L09001 (2007)

7. A. Gautreau, A. Barrat, M. Barthelemy, Global disease spread: statistics and estimation of arrival times. J. Theor. Biol. **251**(3), 509–522 (2008)

8. E.J. Gumbel, Les valeurs extrêmes des distributions statistiques. Annales de l'Institut Henri Poincaré **5**(2), 115–158 (1935)

9. F. Iannelli, A. Koher, D. Brockmann, P. Hövel, I.M. Sokolov, Effective distances for epidemics spreading on complex networks. Phys. Rev. E **95**(1), 012313 (2017)

10. J.G. Kemeny, J.L. Snell, *Finite Markov Chains*, vol. 356 (van Nostrand Princeton, Princeton, 1960)

11. G. Parisi. *Statistical Field Theory*. Addison-Wesley Pub. Co. (1988)

12. M.E. Peskin, D.V. Schroeder, *Quantum Field Theory* (Perseus Books Reading, Massachusetts, 1995)

13. L.G. Valiant, The complexity of enumeration and reliability problems. SIAM J. Comput. **8**(3), 410–421 (1979)

14. J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena* (Oxford University Press, 2002)