SIAM J. APPL. MATH.
Vol. 83, No. 1, pp. 194-224

FRONT PROPAGATION AND ARRIVAL TIMES IN NETWORKS WITH APPLICATION TO NEURODEGENERATIVE DISEASES*

PRAMA PUTRA†, HADRIEN OLIVERI†, TRAVIS THOMPSON‡, AND ALAIN GORIELY†

Abstract. Many physical, epidemiological, or physiological dynamical processes on networks support front-like propagation, where an initial localized perturbation grows and systematically invades all nodes in the network. A key problem is then to extract estimates for the dynamics. In particular, if a single node is seeded at a small concentration, when will other nodes reach the same initial concentration? Here, motivated by the study of toxic protein propagation in neurodegenerative diseases, we present and compare three different estimates for the arrival time in order of increasing analytical complexity: the linear arrival time, obtained by linearizing the underlying dynamical system; the Lambert time, obtained by considering the interaction of pairs of nodes; and the nonlinear arrival time, obtained by asymptotic techniques. We use the classic Fisher–Kolmogorov–Petrovsky–Piskunov equation as a paradigm for the dynamics and show that each method provides different insights but consistent time estimates. Further, we show that the nonlinear asymptotic method also gives an approximate solution, valid in the entire domain, and the correct ordering of arrival regions over large regions of parameters and initial conditions.

Key words. networks, fronts, connectome, neurodegenerative diseases

MSC codes. 92C42, 05C82, 92B20

DOI. 10.1137/21M1467547

1. Introduction.

We live in a connected world where people, goods, information, and diseases travel from one region to the next. In the years of COVID-19, a particularly dramatic example of this propagation phenomenon is, of course, the transmission of the coronavirus from a single seed location (Wuhan, China), to the rest of the world through an intricate network of local and global travel routes [1]. A similar phenomenon also appears in a much smaller system, the human brain, where some toxic proteins associated with neurodegenerative diseases, like Alzheimer's or Parkinson's, are believed to originate in a single region [2] and are transported to the rest of the brain through the so-called *structural connectome*, the network of axonal pathways connecting different regions of the brain [3, 4].

From a mathematical perspective, both phenomena can be understood as the propagation of an autocatalytic process on a network, and the main question is to understand its overall dynamics [5, 6]. For instance, if a process, such as a disease, starts at a seed location, how long will it take to appear at other locations and then develop through a full-scale invasion to produce a global pandemic for a disease, or dementia for the brain? The *arrival-time problem* consists in determining the

* Received by the editors December 23, 2021; accepted for publication (in revised form) November 3, 2022; published electronically February 22, 2023.
https://doi.org/10.1137/21M1467547
Funding: The work of the second, third, and fourth authors was supported by the Engineering and Physical Sciences Research Council grant EP/R020205/1 to the fourth author. The work of the third author was partially supported by the John Fell Oxford University Press Research Fund grant 000872 (project code BKD00160).

† Mathematical Institute, University of Oxford, Woodstock Road, Oxford, OX2 6GG, UK (prama.putra@maths.ox.ac.uk, hadrien.oliveri@maths.ox.ac.uk, alain.goriely@maths.ox.ac.uk).

‡ Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409 USA (travis.thompson@ttu.edu).

time it takes for a quantity of interest to reach a certain level at each node. It has been shown that the scaling of arrival times greatly depends on the coupling dynamics between nodes [7]. Here, we restrict our attention to the important case of linear coupling between nodes, typically given by the graph Laplacian, which is used both for the progression of diseases in global networks [8, 9] and in the modeling of neurodegenerative diseases [10].

1.1. Previous work. There has been a great deal of attention dedicated to the problem of arrival times, especially within the context of epidemiology. There are essentially two different approaches from a modeling perspective: The process is either seen as a stochastic random walk problem on a graph, with some form of autocatalytic expansion at a node [11], or as the discretization of a continuum reaction-diffusion process on the network. While many authors view the stochastic process as the ground truth and the continuum perspective as a mean-field approximation [12, 13, 14], we place ourselves at the continuum level and do not assume that it is driven by a stochastic process [15, 16]. The central question is then to obtain meaningful estimates of the dynamics of the system, rather than justifying the use of a mean-field approach. Therefore, we place ourselves at the continuum deterministic level and do not address further similar models based on stochastic processes.

Within the framework of continuous processes, there are three different approaches to obtain estimates for the arrival time in the case of a nonlinear process on a network. First, one can simplify the topology of the problem and consider the equivalent front propagation on a homogeneous tree or a complete graph. Then, by using a discrete version of the continuous front problem that arises in nonlinear parabolic equations, one can obtain a velocity of propagation [17, 18]. While the mathematical theory for these systems is appealing and has led to many interesting results in their own rights [19], it only applies to networks with very particular structures. However, for the systems we are interested in, the network topology is particularly important and has a strong influence on the arrival times. Hence, we will not consider such methods here.

A second approach consists in studying the linearized system around an equilibrium solution. In this case, the concentration at each node can be obtained as the solution of this linear system for a given initial condition. From this solution, the arrival times are given as the solution of a transcendental equation at each node [20]. This approach is particularly good for small initial concentrations and large diffusion as the linearization provides a decent estimate of the solution. Based on that approach, a suitable ansatz for the nonlinear behavior can be introduced to improve the arrival time estimates [21]. However, the fact that the arrival time is given by a numerical solution of a transcendental equation does not provide much insight into the dynamics. Moreover, without a good prior guess, this solution may be difficult to obtain.

The third approach consists in looking at the propagation between two nodes while neglecting all other connections and, again, linearizing this system. We can then extract an estimate of the time it takes for the second node to reach the same level as the initial condition in the seed node [12]. This arrival time provides a notion of distance between nodes [22], and the arrival time between two distant nodes can then be approximated as the shortest path between these nodes according to this notion of *effective distance* [22]. An improvement based on the contribution due to multiple paths can also be obtained at the cost of extra computations [13]. This

notion of distance on a network is particularly powerful as it gives the network an extra structure from which many estimates can be obtained.

At this point, it is important to reflect on why we need an arrival-time estimate in the first place. If the exercise is to obtain a numerical value, then a direct simulation of the system of nonlinear ordinary differential equations is straightforward to implement and provides an easily computable solution to the problem, even for very large systems. Typically, the solutions obtained by solving transcendental equations in the linear cases, or by using the effective distance over multiple paths, are not as precise and cannot be obtained as easily, in terms of both implementation and computational cost. Hence, the goal here is not to obtain a method to compute the arrival times, as this problem can be solved numerically, but to obtain meaningful estimates and approximations. There are two main reasons why approximations and estimates may be valuable. First, we want to answer basic questions in terms of the parameters: How long does it take for the first seed to reach neighboring nodes? How does the first phase of the process depend on the parameters or topology of the system? Once the initial invasion takes place, can we estimate the velocity of invasion? How long does it take for the system to be fully invaded? Second, we may need a decent analytical approximation of the solutions themselves so that it can be used to understand other coupled processes. We will show that both goals can be achieved by using a combination of approximate distances and asymptotics.

1.2. Fisher-KPP on networks. The simplest model for an invasion process on an undirected connected network is a discrete version of the celebrated Fisher–Kolmogorov–Petrovsky–Piskunov (Fisher-KPP) reaction-diffusion equation [23]. Consider a connected and undirected network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a collection of N nodes $\mathcal{V}$, representing regions of interest, and edges $\mathcal{E}$ representing connections between these regions. If $p_i(t)$ is the quantity of interest in a region i, modeled as a node in a network, and evolving with time t, then on this network, the dynamics is governed by a system of N ordinary differential equations with N initial conditions:

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\rho \sum_{j=1}^{N} L_{ij} p_j + \alpha p_i(1-p_i), \quad i = 1, \ldots, N, \tag{1}$$

$$p_i(0) = q_i, \quad i = 1, \ldots, N, \tag{2}$$

where ρ and α are positive and respectively depict the diffusion constant and growth constant, and $q_i \in [0,1]$ for all i. The matrix $\mathbf{L}$ is the symmetric *graph Laplacian* [24] obtained from the symmetric *weighted-adjacency matrix* $\mathbf{A}$ as

$$L_{ij} = -A_{ij} + \delta_{ij} \sum_{k=1}^{N} A_{ik}, \tag{3}$$

where $\delta_{ij} = 1$ if $i = j$, and 0 otherwise. The graph Laplacian models diffusion processes on a network and encodes the connection between regions: $L_{ij} = 0$ if there is no connection between different nodes i and j. For comparison between different networks, we further assume that the graph Laplacian has been rescaled so that $\max_{i,j}\{A_{ij}\} = -\min_{i,j}\{L_{ij}\} = 1$.

Here, we are using the so-called *standard graph Laplacian* in contrast to the many normalized versions of the Laplacian which are often preferred in the mathematical study of graphs [25]. Indeed, as discussed in [26], our choice of Laplacian is the only

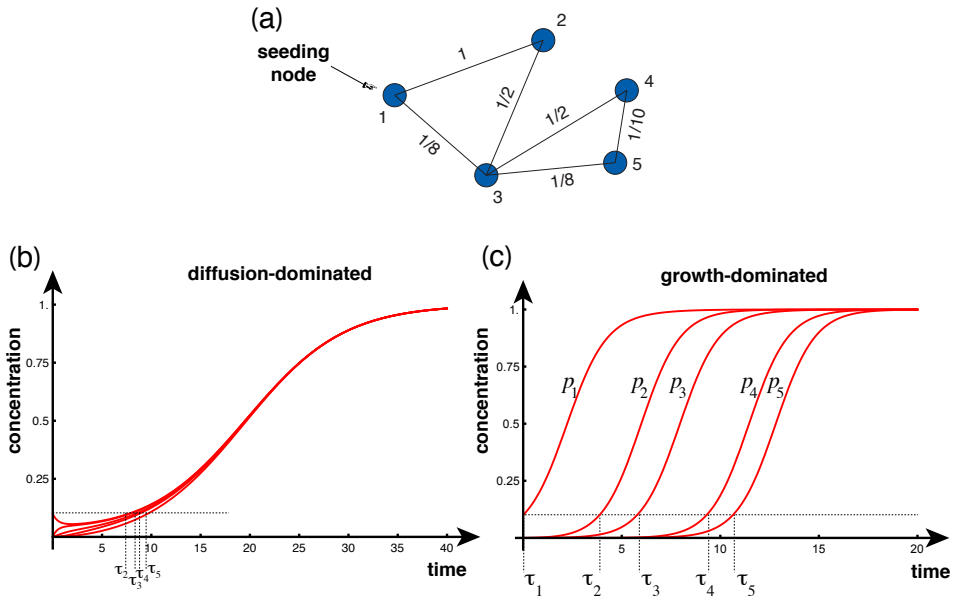

FIG. 1. *Example of dynamics on a* 5*-node network* (a). *Initially, only the first node is seeded* ($p_1(0) = 1/10$). *In the diffusion-dominated case* ((b): $\alpha = 1/5, \rho = 1$), *all concentrations quickly approach* $1/10$ *before experiencing the effect of the growing exponential in unison. For the growth-dominated case, after growth at the first node, the second node is invaded and then all other nodes are subsequently invaded in a front-like progression* ((c): $\alpha = 1, \rho = 1/100$).

one that, in the absence of a reaction term, satisfies the two fundamental properties required to model diffusion: overall mass conservation, and the Fick's property that no transport takes place on the network if the concentrations at each node are equal.

For the *canonical arrival-time problem*, the system is seeded at a single node s so that $q_s = \beta$ and $q_i = 0$ for $i \neq s$. A simple problem is to obtain the value of the *arrival times* $\tau_i = \tau_i(\mu)$ such that $p_i(\tau_i) = \mu$ (with $\beta \leq \mu < 1$ so that τ_i is finite and $\tau_i \geq 0$ with the equality $\tau_i = 0$ attained only if $i = s$ and $\mu = \beta$). Unless otherwise stated, we choose $\mu = \beta$; in this case, the problem is to obtain the times at which other nodes reach the same initial value as the seeding node. The *generalized arrival-time problem* consists of a system with an arbitrary initial condition. Then, letting $\beta = \max_i\{q_i\} > 0$, the corresponding question is to obtain the value of the arrival times $\tau_i = \tau_i(\mu)$ such that $p_i(\tau_i) = \mu$ (with $\beta \leq \mu < 1$ so that $\tau_i \geq 0$ for all i).

Since the network is undirected and connected, the system only supports two equilibrium states when α and ρ are nonvanishing: the *healthy state* defined by $p_i = 0$ for all i, and the *toxic state* defined by $p_i = 1$ for all i. The healthy state is unstable, while the toxic state is stable. The dynamics, in the absence of transport ($\rho = 0$), locally follows a logistic equation with a sigmoid solution. Hence, for all values $0 < \beta < 1$ of the seed, the system will evolve toward the toxic state, which implies that the arrival times τ_i are all positive and finite.

By construction, since the graph is connected, the graph Laplacian has only a single zero eigenvalue. The dynamics is therefore mostly controlled by λ_2, the smallest nonzero eigenvalue of $\mathbf{L}$. Indeed, we can define two main regimes of interest for this system, as shown in Figure 1: the *diffusion-dominated regime*, defined by $\rho\lambda_2/\alpha \gg 1$, and the *growth-dominated regime*, for which $\rho\lambda_2/\alpha \ll 1$.

As a guiding example, consider the 5-node network shown in Figure 1a with weighted adjacency matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1/8 & 0 & 0 \\ 1 & 0 & 1/2 & 0 & 0 \\ 1/8 & 1/2 & 0 & 1/2 & 1/8 \\ 0 & 0 & 1/2 & 0 & 1/10 \\ 0 & 0 & 1/8 & 1/10 & 0 \end{bmatrix}. \tag{4}$$

The graph Laplacian for this network, obtained from (3), has smallest nonzero eigenvalue $\lambda_2 \approx 0.228$.

Intuitively, one expects that in the diffusion-dominated regime, shown in Figure 1b, the concentration will rapidly become uniform over the network, as predicted by a purely diffusive mechanism. Then, a local slow exponential growth will take place more or less simultaneously at all nodes.

The growth-dominated regime is somewhat more interesting from a dynamical point of view. In this case, the concentration at the seeded node will increase while diffusion takes place. The effect of the diffusion term is then to seed other connected regions in which the concentration will then also increase, as shown in Figure 1c. In this case, the system is systematically invaded, node per node, and the question is to understand the dynamics of this cascade by using arrival times as proverbial road markers.

2. A linear analysis.

2.1. General method. The first naive approach to the arrival-time problem is to linearize the system around the healthy state, i.e., $p_i = 0$ for all i, and integrate it. In this case, the system reduces to

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\rho \sum_{j=1}^{N} L_{ij} p_j + \alpha p_i, \quad i = 1, \ldots, N, \tag{5}$$

$$p_i(0) = q_i, \quad i = 1, \ldots, N. \tag{6}$$

A compact solution to this problem is obtained in terms of matrix exponentials. Let $\mathbf{p}(t) = (p_1(t), \ldots, p_N(t))$, $\mathbf{q} = (q_1, \ldots, q_N)$, and $\mathbf{M} = -\rho\mathbf{L} + \alpha\mathbf{1}$, where $\mathbf{1}$ denotes the identity matrix. Then, the solution is given by

$$\mathbf{p}(t) = e^{t\mathbf{M}}\mathbf{q}. \tag{7}$$

Since $\mathbf{L}$ is symmetric, there exists a complete set of orthonormal eigenvectors $\mathbf{v}^{(i)} \in \mathbb{R}^N$ associated with the ordered (but not necessarily distinct) real eigenvalues λ_i, $i \in \{1, \ldots, N\}$, so that

$$\mathbf{L}\mathbf{v}^{(i)} = \lambda_i \mathbf{v}^{(i)}, \quad i = 1, \ldots, N. \tag{8}$$

Further, by construction, we have $\lambda_1 = 0$ with $\mathbf{v}^{(1)} = (1, \ldots, 1)$ and, since the network is connected, $\lambda_i > 0$ for all $i > 1$ [25]. Then, the linear solution can be written explicitly as

$$\mathbf{p}(t) = e^{\alpha t} \sum_{i=1}^{N} \left(\mathbf{q} \cdot \mathbf{v}^{(i)} e^{-\rho\lambda_i t} \right) \mathbf{v}^{(i)}, \tag{9}$$

and the *general linear arrival time* $\widehat{\tau}_j$ is given by the solution of the equation

$$\mu = e^{\alpha\widehat{\tau}_j} \sum_{i=1}^{N} \left(\mathbf{q} \cdot \mathbf{v}^{(i)} e^{-\rho\lambda_i \widehat{\tau}_j} \right) \mathbf{v}_j^{(i)}. \tag{10}$$

If we restrict our attention to the *canonical linear arrival time*, we can further simplify the system by using $q_r = \beta\delta_{rs}$ to obtain

$$\mu/\beta = e^{\alpha\widehat{\tau}_j} \sum_{i=1}^{N} \left(\mathbf{v}_j^{(i)} \mathbf{v}_s^{(i)} e^{-\rho\lambda_i \widehat{\tau}_j} \right). \tag{11}$$

A comparison of the linear solutions with the numerical solutions for our 5-node network is given in Figure 2b, c. A few general features of the method emerge. First, the approximation given by the linear arrival times improves as $\beta = \mu$ decreases. Indeed, as a consequence of the stable manifold theorem, there exists an open set in phase space $\mathbf{p} \in \mathbb{R}^N$ such that the linear solution is a faithful approximation of the solution. Therefore, as $\mu \to 0$ with $\mu > 0$, this approximation converges to the exact value as the nonlinear terms become negligible. This behavior is illustrated in Figure 2c where we compute the error made as a function of the seed size:

$$\mathcal{E}_{\text{lin}}(\mu) = \sum_{i=1}^{N} |\widehat{\tau}_i(\mu) - \tau_i(\mu)|. \tag{12}$$

Second, since the nonlinear terms of (1) act to reduce the value of the concentration in time, and (5) neglects these terms, the solutions of the linear equations are always strictly larger than the actual concentrations (apart from the zero arrival time at the seeding node). Hence, the linear estimates provide a strict lower bound for the arrival times.

Third, to compute the linear arrival times we need to solve N individual transcendental equations of the form (10), one for each arrival time. Solving these equations can prove to be numerically challenging without a decent first guess, a problem that we address in section 3.

Fourth, this method is often overlooked in favor of more involved graph-based measures based on the superposition of multiple paths. However, the exponential matrix solution (7) naturally combines the sum over all paths, due to the properties of the graph Laplacian, as expected from a diffusion operator. Therefore, the linear arrival times constitute a robust universal method valid for small seed concentrations to obtain arrival times, as further demonstrated in sections 3 and 4.

2.2. Diffusion-dominated propagation. The linear analysis is sufficient to understand fully the diffusion-dominated case. Indeed, if diffusion dominates, then the early dynamics is governed by the slowest dynamics in (11), which is the term involving $\exp(-\rho\lambda_2 t)$. Therefore on a typical time scale $t \sim 1/(\rho\lambda_2)$, the concentration at each node tends to the value μ/N. On larger time scales, the effect of the growing exponential is felt and each node behaves, in unison, as a single node subject to the dynamics

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \alpha p(1-p), \quad p(0) = \mu/N. \tag{13}$$

The solution of this equation is

$$p(t) = \frac{\mu e^{\alpha t}}{N - \mu + \mu e^{\alpha t}}, \tag{14}$$

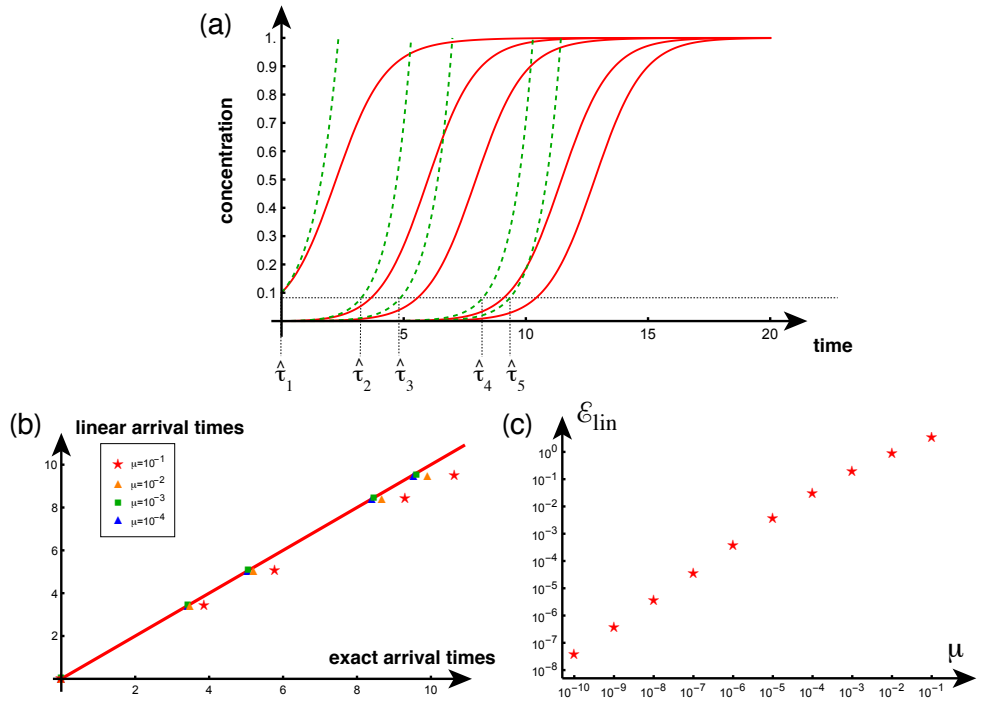

FIG. 2. *Comparison between the exact (solid, red) and linear (dashed, green) solutions for the* 5*-node network* ((a)*:* $\alpha = 1, \rho = 1/100, \beta = 1/10$*). Exact arrival times versus linear arrival times for various initial seeding at node* 1 *(and* $\beta = \mu$*)* (b)*, and error between exact and linear arrival times* (c) *as a function of initial seeding shows good convergence, as expected.*

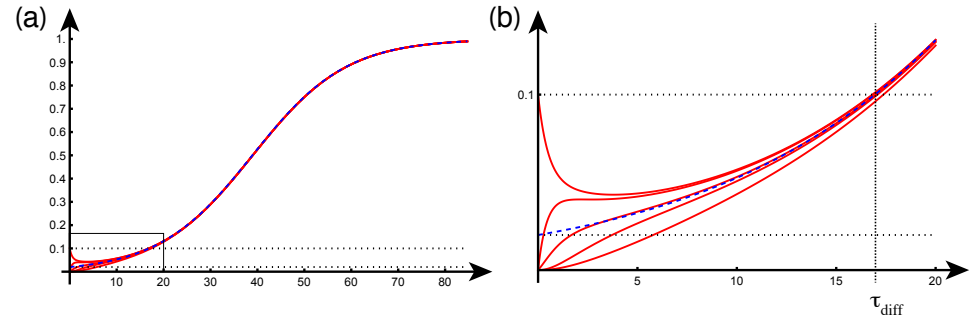

FIG. 3. *Diffusion-dominated case for the* 5*-node network with* $\alpha = 1/10$, $\rho = 1$, $p_1(0) = \mu = 1/10$. *We see that the dynamics of a single node with initial condition* $p(0) = \mu/N$ *(dashed curve) provides an excellent approximation of the entire solution* (a) *and, looking at the close-up in the gray rectangle, for the arrival time as well* (b).

with an arrival time at all nodes given by

$$\tau_{\text{diff}} = \frac{1}{\alpha} \log\left(\frac{N-\mu}{1-\mu}\right). \tag{15}$$

We see in Figure 3 that both estimates give an excellent approximation for the overall behavior of the system and the arrival time.

There are two interesting aspects to this dynamics. First, in the diffusion-dominated case, after an early period, the concentration increases at each node

simultaneously. Hence there is no clear distinction between the different curves, as small changes in the initial conditions can change the order at which each node reaches the critical threshold. Hence, we do not expect to see a well-established ordering of the nodes by arrival times. Second, we see that the concentration at the seeding node decreases initially through diffusion. From the general equations, the slope of the concentration at the seeding node and at the initial time $t = 0$ is given by

$$\sigma = (-\rho L_{ss} + \alpha(1-\mu))\mu. \tag{16}$$

Therefore, the concentration at the seeding node will initially decrease if $\rho L_{ss} > \alpha(1-\mu)$, which will occur if the diffusion constant is large enough. The critical threshold $\rho L_{ss} = \alpha(1-\mu)$ provides another key criterion for the initial dynamics. We note that the diffusion-dominated case is a good example of a general principle for this type of system: if the dynamics between nodes is much faster than the dynamics at the nodes, the overall time-dynamics does not depend on the structural property of the network [27]. Conversely, this rapid diffusion process implies that the structure of the network cannot be obtained given the dynamics of the system as it quickly becomes uniform.

For the rest of the paper, we will focus on the richer growth-dominated case, for which we assume, without loss of generality, that α is of order one and ρ is a small parameter.

3. The Lambert distance. A second approach is to define a notion of distance on edges, by considering the time it takes for an initial seed to propagate to a neighboring node, while ignoring all other nodes in the network. Consider two nodes i and j connected by an edge with weight $A_{ij} \neq 0$. Neglecting all other nodes in the network, we consider the linear approximation (5) with surrogate graph Laplacian, corresponding to the 2-node subnetwork of node i and j, given by

$$\mathbf{L}_{ij} = \begin{bmatrix} A_{ij} & -A_{ij} \\ -A_{ij} & A_{ij} \end{bmatrix}.$$

Using the eigenvectors of $\mathbf{L}_{ij}$, i.e., $\lambda_1 = 0$ and $\lambda_2 = 2A_{ij}$, and (11) with $\mu = \beta$ yields an equation for the arrival time at node j, from an initial seeding at node i (or vice versa), given by

$$e^{\alpha t}\left(1 - e^{-2\rho A_{ij} t}\right) = 1. \tag{17}$$

Expanding for small ρ, we obtain

$$\rho A_{ij} t e^{\alpha t} = 1. \tag{18}$$

The solution of this equation can be expressed in terms of the Lambert function W_0 (defined so that the real solution of $te^t = z$ is $t = W_0(z)$):

$$t_{ij} = \frac{1}{\alpha} W_0\left(\frac{\alpha}{\rho A_{ij}}\right), \qquad i \neq j. \tag{19}$$

We note that since the network is undirected, we have $A_{ij} = A_{ji}$ which implies that $t_{ij} = t_{ji}$.

We refer to t_{ij} as the *Lambert edge distance*. From this pairwise distance between neighboring nodes, we define the *Lambert distance* W_{ij} as the shortest path with

respect to the Lambert edge distance between two nodes i and j. Explicitly, let $\Gamma_{ij} = (\gamma_0, \gamma_1, \ldots, \gamma_n)$ with $\gamma_0 = i$ and $\gamma_n = j$ be this shortest path; then

$$W_{ij} = \begin{cases} \displaystyle\sum_{k=0}^{n-1} t_{\gamma_k,\gamma_{k+1}} = \frac{1}{\alpha} \sum_{k=0}^{n-1} W_0\left(\frac{\alpha}{\rho A_{\gamma_k,\gamma_{k+1}}}\right), & i \neq j, \\ 0, & i = j, \end{cases} \tag{20}$$

which defines the *Lambert distance matrix* $\mathbf{W}$. By construction, this distance is a metric on the network (as $W_{ii} = 0\ \forall i$, $W_{ij} = W_{ji}\ \forall i, j$, and $W_{ij} \leq W_{ik} + W_{kj}\ \forall i, j, k$).

The Lambert distance provides a useful estimate for the arrival times. To take into account the fact that, in general, the critical concentration may be different from the initial concentration, $\mu \neq \beta$, we define the *self-time* t_{ii} to be the time at which a local initial concentration β, at node i, reaches μ in the absence of any connection. That is, the time t_{ii} at which the solution to

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \alpha p(1-p), \quad p(0) = \beta,$$

satisfies $p(t_{ii}) = \mu$. The solution of this problem is

$$t_{ii} = \frac{1}{\alpha} \log\left(\frac{(1-\beta)\mu}{(1-\mu)\beta}\right). \tag{21}$$

In the canonical case of a single seed at node i, the *Lambert arrival times*, at every node j, are defined by

$$\tilde{\tau}_{ij} = t_{ii} + W_{ij}. \tag{22}$$

Since both W_{ij} and t_{ij} are symmetric, it follows that $\tilde{\tau}_{ij} = \tilde{\tau}_{ji}$. It is useful to note that the term W_{ij}, in (22), is computed by (20), which implicitly assumes that $\mu = \beta$ via (17)–(19), and that the self-time, t_{ii} of (21), vanishes when $\mu = \beta$. The Lambert arrival times (22) can then be understood as the sum of the time it takes for node j to reach the same concentration as the seed node i and the time for node j to reach concentration μ starting with concentration β. Equivalently, it can also be interpreted as the sum of the time it takes for node i to reach concentration μ and the time it takes for the concentration μ at node i to reach node j.

This estimate misses three contributions, which prevents it from being an upper or lower bound. First, any node before the target node is connected to other nodes (including previous nodes). Hence, the quantity of interest is diluted away from the target node, which results in an underestimation of the arrival time. Second, the target node may also receive contributions from other paths, resulting conversely in overestimated arrival time. Third, the estimate is based on a linearization of the equations; hence it misses saturating effects that increase the arrival time. Thus, unlike the linear arrival times, the Lambert arrival times (22) do not provide a general bound, neither upper or lower, for the true arrival times.

Yet, as shown next, the Lambert method provides a remarkably good estimate despite its simplicity. It has a few notable features. First, the entries of the Lambert distance matrix (20) are independent of the initial concentration as the initial concentration acts only to perturb the Lambert arrival time (22) via the self-arrival (21). The Lambert distance, between two distinct nodes, is therefore an intrinsic property

of the system. It implies that the initial seed concentration does not appreciably impact the overall propagation. Second, the Lambert method highlights the role of the diffusion in the arrival-time problem and, in particular, shows that diffusion has only a weak influence on this quantity. For instance, for $\rho/(\alpha W_{ij}) \approx 0.01$, an increase of diffusion by a factor 10 only changes the arrival times by a factor 2. Third, the Lambert arrival times can be used as a natural first guess to find the linear times and greatly improve the convergence of root-finding methods applied to (11).

From the Lambert arrival times, we can obtain simple expressions for the front solution. This is done by assuming that the dynamics after the arrival time is controlled by the local dynamics so that, after this time, the effect of the graph Laplacian is neglected. More specifically, assume that node i is chosen as the seeding location, with initial seed $p_i(0) = \mu$; then the dynamics at node j is given by

$$\frac{\mathrm{d}\tilde{p}_j}{\mathrm{d}t} = \alpha \tilde{p}_j(1 - \tilde{p}_j), \quad \tilde{p}_j(\tilde{\tau}_{ij}) = \mu, \tag{23}$$

where $\tilde{\tau}_{ij}$ is the Lambert arrival time (22). The *Lambert solution*

$$\tilde{p}_j = \frac{\mu e^{\alpha t}}{\mu e^{\alpha t} - (\mu - 1)e^{\alpha \tilde{\tau}_{ij}}}, \quad j = 1, \ldots, N, \tag{24}$$

is therefore an approximation of (1).

The Lambert edge distance (19), Lambert distance (20), and Lambert solution (24) can be readily illustrated using the exemplar 5-node network of Figure 1. The matrices corresponding to the Lambert edge distance and Lambert distance, respectively, are

$$[t_{ij}] = \begin{bmatrix} \infty & 3.39 & 5.06 & \infty & \infty \\ 3.39 & \infty & 3.93 & \infty & \infty \\ 5.06 & 3.93 & \infty & 3.93 & 5.06 \\ \infty & \infty & 3.93 & \infty & 5.25 \\ \infty & \infty & 5.06 & 5.25 & \infty \end{bmatrix}, \tag{25}$$

$$W = \begin{bmatrix} 0.00 & 3.39 & 5.06 & 8.99 & 10.10 \\ 3.39 & 0.00 & 3.93 & 7.86 & 8.99 \\ 5.06 & 3.93 & 0.00 & 3.93 & 5.06 \\ 8.99 & 7.86 & 3.93 & 0.00 & 5.25 \\ 10.10 & 8.99 & 5.06 & 5.25 & 0.00 \end{bmatrix} \tag{26}$$

where we have used the symbol ∞ to characterize pairs of nodes that are not connected by an edge (and hence have an infinite Lambert edge distance). An example of the corresponding Lambert solutions and the arrival times is given in Figure 4. The Lambert solutions, and Lambert arrival times, are seen to be an improvement over both the linear solutions, and arrival times, and for capturing the dynamics over the entire domain. In Appendix D, we compare the Lambert distance with the so-called effective distance that is popular to study arrival times.

Finally, we introduce the notion of *step distance*. The step distance is a vector, $\boldsymbol{\gamma}$, whose jth entry is the number of edges along the shortest path from the node of initial seeding to the node j. For the 5-node network example of Figure 1, with seeding at node 1 we have $\boldsymbol{\gamma} = (0, 1, 1, 2, 2)$. Thus, for example, this particular $\boldsymbol{\gamma}$ indicates that the shortest path, with respect to the Lambert metric established by $\mathbf{W}$, between seed node 1 and target node 3 is composed of two edges.

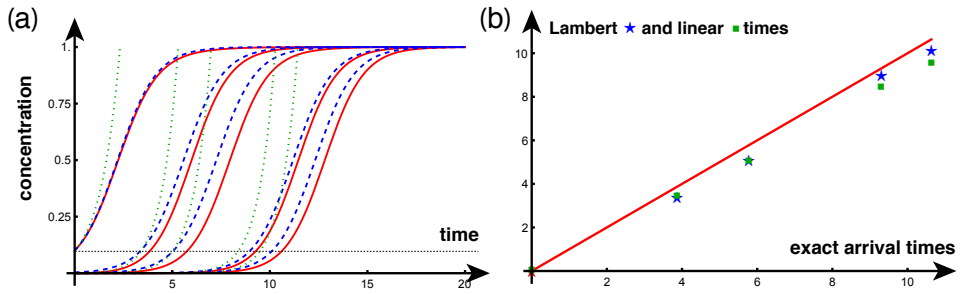

FIG. 4. (a) *Exact (red, solid), Lambert (blue, dashed), and linear (green, dotted) solutions with* $\alpha = 1$, $\rho = 1/100$, $p_1(0) = \mu = \beta = 1/10$. (b) *Lambert arrival times (blue stars) compared to linear arrival times (green squares).*

3.1. Multiple seeding locations. The Lambert distance assumes that a single node i has been seeded and gives the arrival times W_{ij} at node j. In the case where multiple seeding locations are given at time 0, the method can be easily adapted. Indeed, assume that a general initial condition $\mathbf{p}(0) = \mathbf{q}$ is given, with $0 \le \max_i\{q_i\} < \mu \ll 1$. The current task is to derive a closed-form expression for an approximate arrival time $\tilde{t}_j$, in a target node j, such that $p_j(\tilde{t}_j) = \mu$. Fix a target node j and consider a seed node, i, such that $0 < q_i$. Following the previous analysis, we use (22), with $\beta = q_i$, to obtain

$$\tau_{ij} = \frac{1}{\alpha}\log\left(\frac{(1-q_i)\mu}{(1-\mu)q_i}\right) + W_{ij}, \tag{27}$$

which estimates the arrival time at node j due to the seed at node i. Now, we construct an approximation of the contribution to the concentration at node j due to the seeding at node i by

$$\tilde{p}_j^{(i)}(t) = \mu\frac{t\exp(\alpha t)}{\tau_{ij}\exp(\alpha\tau_{ij})}. \tag{28}$$

Summing over the partial contributions from the multiple seeding sites yields an approximation of the total concentration at node j given by

$$\tilde{p}_j(t) = \sum_{i=1}^{N}\tilde{p}_j^{(i)}(t) = \mu t\exp(\alpha t)\sum_{i=1}^{N}\frac{1}{\tau_{ij}}\exp(-\alpha\tau_{ij}), \tag{29}$$

which reaches the level μ at time

$$\tilde{t}_j(\mu) = \frac{1}{\alpha}W_0\left(\frac{\alpha}{\sum_{i=1}^{N}\frac{1}{\tau_{ij}}\exp(-\alpha\tau_{ij})}\right). \tag{30}$$

Just as in the case of a single seeding site, the estimate (30) can be used as an initial guess to compute the linear arrival times for the case of multiple seeding locations.

4. Nonlinear approximations. The estimates obtained so far are useful, but they do not fully exploit all the information that can be obtained from the nonlinear system. In particular, these methods do not take into account the coupling between the nonlinear sigmoid response encapsulated in the nonlinear term and the diffusion

from nearby nodes given by the graph Laplacian. In the growth-dominated case, we have a natural small parameter that can be used for asymptotic computation.

Assume, by a simple rescaling in time and without loss of generality, that α is of order one and that $\varepsilon = \rho \ll 1$ is a small parameter. A naive asymptotic method that systematically expands all solutions in terms of ε, and solves the resulting equations order by order, fails to provide any useful results as it only provides a small correction of the exponential behavior which is unbounded in time.

To circumvent this difficulty, we present an alternative asymptotic method that captures important features of the solutions and improves the computation of the arrival times. We consider the canonical case and start with the solution with $\varepsilon = 0$ given by

$$\tilde{p}_i = \frac{K_i e^{\alpha t}}{1 - K_i\,(e^{\alpha t} - 1)}, \quad i = 1, \ldots, N. \tag{31}$$

At this point, we resist the urge to set the arbitrary constant from the initial conditions. Rather, we use a nonlinear version of the method of variation of constants[1]

$$K_i(t) = A_i + \varepsilon^{\gamma_i} B_i(t), \tag{32}$$

where γ_i is the step distance between the seed node and node i as defined in section 3). With this ansatz, we can solve iteratively for the unknown function B_i for the subsequent subsets of nodes with increasing distance from the seeding node up to the furthest node given by the maximum of $\boldsymbol{\gamma}$.

To lowest order $\mathcal{O}(\varepsilon^0)$, the solution at the seeding node with initial condition $p_s(0) = \beta$ is simply the original unperturbed solution

$$\tilde{p}_s = \frac{\beta e^{\alpha t}}{1 - \beta\,(e^{\alpha t} - 1)}, \tag{33}$$

and $A_i = 0$ for all $i \neq s$. To order $\mathcal{O}(\varepsilon^1)$, representing all nodes i at a distance one from the seeding node, the differential equation for B_i is

$$\frac{\mathrm{d}B_i}{\mathrm{d}t} = -L_{is} e^{-\alpha t} \tilde{p}_s, \quad B_i(0) = 0, \tag{34}$$

that can be easily integrated to obtain

$$B_i(t) = \frac{\beta L_{is}}{\alpha(\beta - 1)} \left(\alpha t - \log\left(\beta\left(e^{\alpha t} - 1\right) + 1\right)\right). \tag{35}$$

The solutions for nodes at distance two from the seeding node can then be obtained iteratively with the solution at nodes located at distance d depending only on the solutions given by nodes at distance $d - 1$. More formally, if we define the matrix

$$M_{ij} = \begin{cases} L_{ij} & \text{if } \gamma_i - \gamma_j = 1, \\ 0 & \text{otherwise} \end{cases} \tag{36}$$

[1]The method used here is a bit unorthodox. However, it can be seen as a more or less straightforward generalization of the well-known method of variation of constants for inhomogeneous linear differential equations. As discussed in [28], this method was first proposed by Lagrange in 1788 [29], almost exactly along the lines used here.

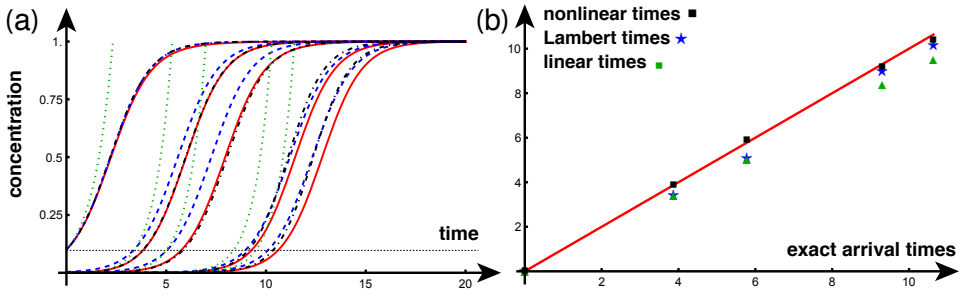

FIG. 5. (a) *Exact (red, solid), Lambert (blue, dashed), linear (green, dotted), and asymptotic (black, dash-dotted) solutions with* $\alpha = 1, \rho = 1/100, p_1(0) = \mu = 1/10$. (b) *Asymptotic arrival times (black squares), versus Lambert arrival times (blue stars), versus linear arrival times (green squares).*

and the variables

$$Z_i = \begin{cases} B_i, & \text{if } i \neq s, \\ e^{-\alpha t}\tilde{p}_s & \text{if } i = s, \end{cases} \tag{37}$$

then, these new variables are solutions of the system

$$\frac{\mathrm{d}Z_i}{\mathrm{d}t} = \sum_{j=1}^{N} M_{ij} Z_j, \qquad Z_i(0) = 0, \qquad i \neq s. \tag{38}$$

The asymptotic solutions, for the 5-node network of Figure 1, are shown in Figure 5 together with the computations of the arrival times.

Remarkably, the solution of (38) can be obtained explicitly in terms of the polylogarithm function Li_d, where d is the distance to the seeding node (for node at distance 1, the polylogarithm reduces to the natural log as found in (35)). Once substituted in (31), these *asymptotic solutions* satisfy both the initial conditions and the asymptotic behavior for long time. They naturally take into account the contributions from all paths of length d rather than just a single path. Hence, they improve the Lambert solution, albeit at the cost of increasingly long expressions better carried through symbolic manipulations in the privacy of one's office. They do not include the contributions due to paths of different lengths or the effect of dilution away from a node. Both of these effects could be, in principle, included in the matrix $\mathbf{M}$ by adding the corresponding couplings and introducing higher-order corrections, but at the cost of an increase in complexity that may not be warranted.

5. Application to neurodegenerative diseases. In this section, we apply the arrival time methods to the study of propagation of misfolded toxic protein species, transported on a structural connectome.

5.1. Modeling toxic protein propagation. We based our modeling approach on the so-called *prion-like hypothesis* of neurodegenerative diseases stating that, much like prion diseases, neurodegenerative diseases are directly related to the systematic expansion of toxic protein aggregates and their transport through the axonal pathways known as the structural connectome [30, 31, 32, 33, 34, 35, 36]. Indeed, it is now well appreciated that neurodegenerative diseases such as amyotrophic lateral sclerosis, Parkinson's, and Alzheimer's are associated with toxic proteins forming large

aggregates. Starting in an initial region, these toxic proteins systematically invade the brain causing neurotoxicity and loss of functions.

From a modeling perspective, there are multiple ways to encode these mechanisms either in a mathematical or a computational model. Early attempts model the spread through a diffusion mechanism on the graph of the structural connectome with great success [10, 37, 38, 39, 40], but these models lack a proper mechanism for autocatalytic expansion. To take this effect into account, computational models have been introduced with specific rules for transport and expansion that also include stochastic effects [41, 42, 43, 44, 45]. These models have shown an excellent match with data and can help for classification purposes. However, they lack the mathematical simplicity that is needed to identify key mechanisms of importance. An alternative approach is to start from first principles and take into account the physical mechanisms of protein aggregation, leading to an infinite set of Smoluchowski-type differential equations for the concentration of oligomers of different sizes [46, 47, 48, 49, 50, 51]. When clearance mechanisms are taken into account, an instability takes place in the system whereby below a critical protein clearance level, monomers start accumulating into larger and larger aggregates [4]. In recent work, we showed that close to that instability, the system is universally described by a transcritical instability for the total mass of toxic proteins [52]. When transport on the structural connectome is added to the model, these equations take the form of the network-Fisher-KPP equations (1). This model was extensively studied and shown to reproduce carrying capacity curves of continuous models [53] and of Smoluchowski-type models [4]. Further, Bayesian analysis was used to validate the models against patient data and extract suitable parameters [54, 55]. Based on this work and the data analysis of [56] for typical expansion time, we use for the small resolution network (see below) a diffusion constant $\rho = 0.01/\text{year}$. The interval for the growth constant is $[0.2, 0.6]$, and we use a growth constant $\alpha = 0.5/\text{year}$.

5.2. Modeling the structural connectome. A *structural connectome* and its weighted adjacency matrix $\mathbf{A}$ can be obtained by tractographic reconstruction of axonal bundles from diffusion tensor magnetic resonance images. The structural network of a patient brain is then defined by a set of vertices reflecting cortical and subcortical regions of interest and a set of edges representing the axonal bundles that connect these regions. The structural networks used here are averaged from 426 patient data graphs that are part of the Budapest Reference Connectome v3.0 [57, 58] and generated using a multiresolution Lausanne anatomical parcellation [59, 60]. The connectome networks are freely available [59] at five levels of resolution, and we will show the main results on the averaged connectome with the lowest ($N = 83$) resolution [61]. The other network sizes will be used for a systematic evaluation of the approximation errors. These networks have edges with *diffusive weights* A_{kj} between nodes k and j, defined as the ratio of the number of axonal fibers n_{ij} between the nodes and the mean fiber length ℓ_{ij} squared in order to capture the main scaling properties of diffusion [26]:

$$A_{ij} = \frac{n_{ij}}{l_{ij}^2}, \quad i, j = 1, \dots, N. \tag{39}$$

With these brain networks in hand, different mathematical models of misfolded protein propagation have been systematically used and validated to study key aspects of neurodegenerative diseases [62, 63, 61]. For instance, Alzheimer's disease is associated with the formation of large aggregates of amyloid-beta and tau proteins. The accumulation of toxic tau proteins is particularly important in understanding the disease as

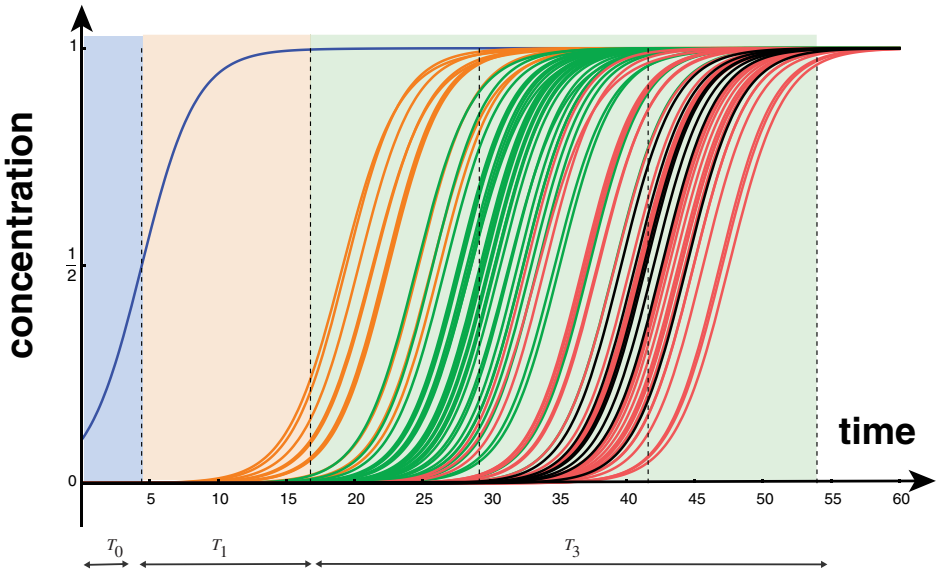

FIG. 6. *Time scale estimates for the $N = 83$ connectome. Following the local growth at the seeding node with a time scale T_0, the first neighborhood is invaded after a time T_1. Successive nodes are invaded progressively with a general time scale depending on the number of steps to the last node. Parameters: $\alpha = 0.5$/year, $\rho = 0.01$/year. Initial condition: $p_{27} = 1/10$ and $p_i = 0$ for all other $i \in \{1, \ldots, N\}$. The colors correspond to γ as defined in the text: the number of steps from the seed according to the Lambert metric (0: blue, 1: orange, 2: green, 3: red, 4: black). (Color figure available online.)*

it is associated with regions of atrophy and cognitive dysfunction. Its progression has two key features of relevance for the present work: the toxic load originates in localized areas of the brain around the entorhinal cortex [23], and a systematic pattern of progression has been identified through postmortem histopathology, suggesting that the progression is in the growth-dominated regime with a small diffusion constant [64, 55].

5.3. Time scales of disease progression. In Figure 6, we show the solution of the Fisher-KPP equation on the small scale connectome with 83 nodes for particular initial conditions and parameter values. We are now in a position to reflect on the different methods and extract from our analysis important features that answer the question we started with: How do local spreading and global propagation depend on the system's parameters? In the growth-dominated case, we see a clear separation of time scales. The *local time scale,*

$$T_{\text{loc}} = 1/\alpha, \tag{40}$$

governs the local increase of any seed. Indeed, an initial concentration $\beta < 1/2$ reaches locally a concentration of $1/2$ at a time

$$T_0 = T_{\text{loc}} \log\left(\frac{1-\beta}{\beta}\right). \tag{41}$$

The second important time scale is the arrival time at a neighboring node. For a given seed at node s, it is given by

$$T_1(s) = \frac{1}{\alpha} W_0\left(\frac{\alpha}{\rho a_s}\right), \quad a_s = \max_i A_{is}, \tag{42}$$

where a_s is the maximum of the sth column vector of matrix $\mathbf{A}$. If the entries of the weighted adjacency matrix, $\mathbf{A}$, are normalized by the maximal entry, then, over all possible seeds, the quantity $\max_i\{A_{is}\}$ is, by construction, equal to one. In this case, we can define, for the entire network, the time scale

$$T_1 = \min_s T_1(s) = \frac{1}{\alpha} W_0\left(\frac{\alpha}{\rho}\right) \tag{43}$$

as the *denial period*. The *denial period* (43) corresponds to a time scale quantifying the initial lag in disease invasion from any epicenter. Before the denial period has lapsed, the disease appears to be located at the source and its propagation to neighboring nodes is not yet measurable. The denial period only depends on local connectivity and does not rely on the global, possibly complex, topological characteristics of the network.

If a seeding site is known, or suspected, the local denial period gives a good estimate of the time before an invasion is detectable in nearby regions. The denial period estimates are particularly interesting in the case of modeling Alzheimer's disease and may provide a window into the early phases of progression for both a typically assumed entorhinal cortex seeding [53, 62, 65] and the potential alternative seeding sites referenced in connected to nonstandard Alzheimer subtypes [66]. Conversely, the global denial period (43) represents an estimate of the minimal window of time necessary for any toxic protein to spread from any seeding location within the brain and is a property of the connectome graph and the choice of weights in the matrix $\mathbf{A}$.

The third time scale corresponds to the rapid succession of invasion fronts as shown in Figure 7. We define, for $i > 1$, $T_i(s) = iT_1(s)$ as the typical scale at which nodes that have values i with respect to the step distance defined in section 3 are affected by an initial seeding in node s. For instance, the green curves in Figure 6 characterize nodes whose shortest path to the seeding node has two edges. We see from Figure 6 that for this network, it is the edge distance that mostly defines the arrival time.

More generally, for a network model over which the system (1) is evolved, the important time scales are T_{loc} for the initial growth, T_1 for the first initial spread, and the *pandemic period* where most of the graph has been invaded:

$$T_{\max} = (\ell + 1)\, T_1, \tag{44}$$

where ℓ is the maximal number of edges among all shortest paths between two nodes in the network defined by $\mathbf{A}$. These time scales are independent of the seeding node and the initial condition. These estimates can be further refined for a given initial seed by replacing T_1 by $T_1(s)$.

5.4. Arrival times and Braak staging. Despite the overall simplicity of the connectome Fisher-KPP model, its dynamics captures well the generalized progression of tau protein neurofibrillary tangles observed in postmortem studies. In particular, histopathological studies have noted that neurofibrillary tangles follow a six-stage sequence [67, 68], called the *six Braak stages* [36]. While generalized progression

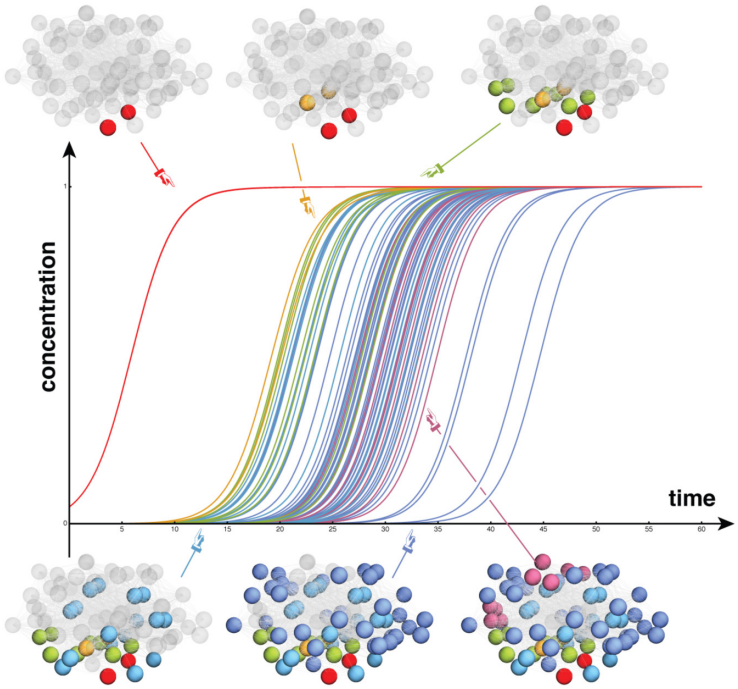

FIG. 7. *Starting at the entorhinal cortex, the dynamics of the network Fisher-KPP equations recovers some of the key aspects of Braak staging:* $N = 83, \alpha = 0.5/\text{year}$, $\rho = 0.01/\text{year}$. *Initial conditions:* $p_{27} = p_{68} = 1/20$ *and* $p_i = 0$ *for all other* $i \in \{1, \dots, N\}$. *Each of the six colors corresponds to one of the cortical regions that compose a Braak stage. For instance, regions that are part of Braak* I *are in red. (Color figure available online.)*

sequences have also been proposed [69, 70], here we limit ourselves to the six Braak regions used in the standardized positron emission tomography AV-1451 data pipeline [71] of the Alzheimer's Disease Neuroimaging Initiative. In line with histopathological observation [67, 68] we expect a progressive ordering of arrival, of tau concentration in different regions, starting in Braak region I corresponding to the entorhinal cortex. Therefore, we solve (1) on a brain connectome with an initial seeding in the bilateral entorhinal cortex. Figure 7 shows the dynamics of an example simulation, with illustrative parameters in the growth-dominated regime, using the lowest resolution $N = 83$ connectome. In this case, the entorhinal seeding sites correspond to nodes 27 and 68.

We apply the three methods for arrival times on this problem for the same parameters as in Figure 7 and compute the approximate linear solutions. In Figure 8, we give the three arrival times and compare, for each Braak region, the exact numerical solution (solid curves), with the approximate asymptotic solutions (dashed curves). To visualize the results and the quality of the approximations, we average the concentration for each Braak region (as the sum of the concentration over each node is divided by the number of nodes in that region). We observe first that the approximation provided by the asymptotic solution is excellent on the entire domain and, remarkably, that the model correctly captures the staging pattern observed in Alzheimer's disease. To compare different approximations for networks of different

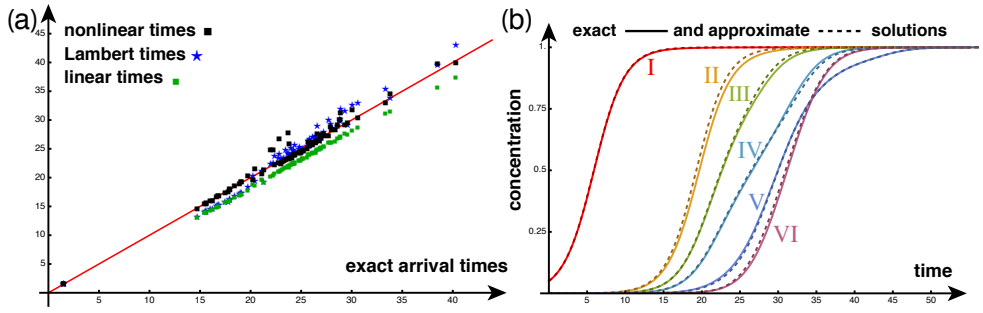

FIG. 8. *Arrival times at $\mu = 1/10$ for $N = 83$ connectome (left) and average concentration (right) in each Braak region (parameters and color coding coincide with Figure 7). Solid curves are the numerical solutions and the dashed curves their approximation from the asymptotic expansion. Initial conditions are chosen so that the total concentration in the entorhinal cortex is $1/10$: $p_{27} = p_{68} = 1/20$ and $p_i = 0$ for all other $i \in \{1, \dots, N\}$. (Color figure available online.)*

sizes under different dynamics, we introduce the relative error, with respect to the average arrival time:

$$\mathcal{E}_{\text{approx}} = \frac{\sum_{i=1}^{N} |\widehat{\tau}_i - \tau_i|}{\sum_{i=1}^{N} \tau_i}, \qquad \text{approx} \in \{\text{linear, Lambert, nonlinear}\}, \tag{45}$$

where τ_i is the exact (numerical) arrival time at node i and $\widehat{\tau}_i$ is one of its approximations given by one of the three methods. The average arrival time for this particular case is 23.6 years, and as expected, the nonlinear solution provides a decent approximation of the solution and the best arrival times with comparative scores (12) given by

$$\mathcal{E}_{\text{linear}} \approx 84 \times 10^{-3}, \quad \mathcal{E}_{\text{Lambert}} \approx 41 \times 10^{-3}, \quad \mathcal{E}_{\text{nonlinear}} \approx 24 \times 10^{-3}. \tag{46}$$

5.5. Data analysis. When recordings are made of tau protein concentration levels in the brain, there is no information about time since onset. Therefore, we cannot directly use the arrival times to learn from these data sets. However, we can use the approximation provided by the asymptotic solutions to extract meaningful information. Here, as an example we study recently published data about tau concentrations at different stages of the disease to obtain estimates on their timing [72]. The Korean data set gives the concentration levels of tau protein in 86 brain regions for the three typical disease stages: normal, amnestic mild cognitive impairment, and Alzheimer's disease (see details about methods and data treatment in [72]).

First we map the 86 brain region atlas to 82 of our nodes (the original atlas does not contain information about the brain stem which is therefore omitted in our study). Second, we assume that the scans of normal individuals provide a baseline for tau protein concentration and off-site binding and we obtain a concentration score by subtracting these values from the mild cognitive impairment and Alzheimer's disease data set at each node and scale the data set so that the maximal value is 1. Doing so, we obtain two data sets given by a vector with 82 entries: $C^{(\text{M})} = \{C_i^{(\text{M})},\ i = 1, \dots, 82\}$ for patients with mild cognitive impairment and $C^{(\text{A})} = \{C_i^{(\text{A})},\ i = 1, \dots, 82\}$ for patients with Alzheimer's disease.

The question is then to find the time from onset more likely to coincide with the data for the two disease stages. Following previous studies [73, 40, 74], we do so

by computing the Pearson correlation coefficients $R^{(\mathrm{M})}(t)$ and $R^{(\mathrm{A})}(t)$ between the data and the asymptotic approximation of the concentrations given by (31)–(38) for all time t and identifying the time at which each coefficient is maximal. To check the validity of our approach, we also run the full numerical model and compute the correlation coefficients against the data as shown in Figure 9. Our method gives a maximum correlation coefficient of about 0.7 for the mild cognitive impairment data at a time around 20.5 years and a coefficient around 0.5 for the Alzheimer's disease data at 26.5 years. Both coefficients are slightly higher but consistent with the coefficients given in [72] for a full aggregation model. These values are consistent with the average arrival time of about 23 years and a progression from mild to severe disease of 6 years. Note that we have not tried to find the model parameters that maximize the correlation but have found that variations in these parameters did not affect notably the value of the correlation coefficients (with values up to 0.8 in the physical range for the growth constant). Model validation and parameter fitting would require a full Bayesian analysis on larger sets of data as performed in [64, 55].

5.6. Braid diagrams and braid surfaces. We have applied each approximation method on the lowest and highest resolution connectomes for particular values and are interested in assessing the performance of our methods systematically through staging analysis in the growth-dominated regime ($\rho/\alpha \ll 1$) as introduced in Putra et al. [26]. The idea is to only consider the ordering of the arrival times rather than their numerical values. Following [26], we can plot stagings, across all possible values of the threshold μ, by using a *braid diagram*. By way of illustration, Figure 10a shows a simple example of a braid diagram for our exemplar 5-node network given in Figure 1. The braid diagram displays for each value of μ, starting at $\mu = \beta = 1/100$, the order at arrival. For $\mu = 1/100$, the staging is $(3, 4, 2, 1, 5)$ indicating that node 3 is the first one at which the concentration reaches $1/100$ followed by node 4, then 2, 1, and 5. For a higher threshold, the ordering changes. For instance, at $\mu = 1/2$, the ordering is $(4, 3, 2, 1, 5)$. A *braid surface* summarizes a parametric family of braid diagrams. To capture this variability visually, we introduce the notion of *braid surface* that allows for representing multiple braid diagrams associated with different values of ρ/α [26]. The idea is to assign a single integer value, in $[1, N!]$, and to assign a color label to that integer. For our example, the braid surface in Figure 10b gives eight possible stagings depending on both threshold and parameters. Braid surfaces are a powerful means of summarizing staging on a general graph and have been used [26] to show that variations in parameters, connectome resolution, tractography methods, and thresholding can cause significant differences in observed Braak staging.

Constructing a braid surface can be a computationally expensive task. Suppose that μ and $\log(\rho/\alpha)$ take values in an interval I. Determining a braid surface, corresponding to this value range, by exact means requires the discretization of I into M and R subintervals, for μ and $\log(\rho/\alpha)$, respectively, and $M \times R$ solutions of the stiff Fisher-KPP system. The arrival times can also be approximated by our methods discussed. In practice, this trades the computational task of solving a stiff system for the task of solving the problems associated to the estimation methods. Of the three methods, the Lambert method is the most computationally attractive. The Lambert method involves computing the Lambert function for each edge in the graph and determining a shortest path; both tasks have highly performant implementations in many common computing software packages. Moreover, edge distances and shortest paths can be reused for all values of μ.

To test the use of the Lambert distance to compute the braid surfaces, we use the *Kendall tau* distance [75] to measure the distance between two stagings as

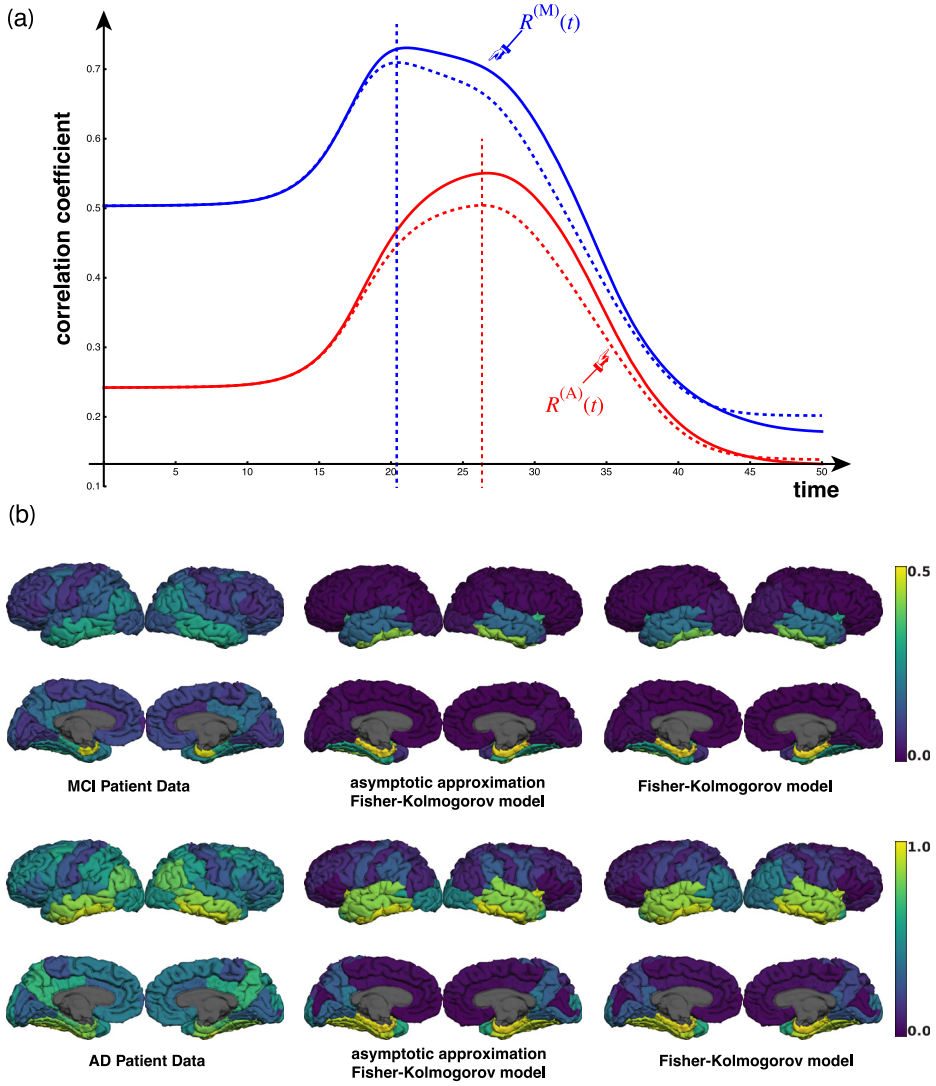

FIG. 9. (a) *Blue: Correlation coefficient* $R^{(A)}(t)$ *between concentration data from patient with mild cognitive impairment and the numerical (solid) and analytical (dashed) output of the network-Fisher-KPP model. Red: Correlation coefficient* $R^{(M)}(t)$ *between concentration data from patient with Alzheimer's disease and the numerical (solid) and analytical (dashed) output of the network-Fisher-KPP model (all model parameters and initial conditions as in Figure* 8*. (Color figure available online.)*

the minimum number of transpositions necessary to transform one staging into the other. For instance, the Kendall tau distance between stagings $(1,2,3)$ and $(1,3,2)$ is $d_K((1,2,3),(1,3,2)) = 1$, since transposing the last two values of $(1,3,2)$ brings the stagings into agreement. Using this notion of distance, we can assess the errors between a braid surface computed explicitly, using (1), and those computed using the linear, Lambert, and nonlinear estimation methods. Figure 11a shows an exact braid surface corresponding to Braak staging on the $N = 83$ node connectome; the yellow color indicates the canonical Braak staging sequence for tau proteins as described by Braak [67, 68]. Figures 11b–d show the Kendall tau distance between the

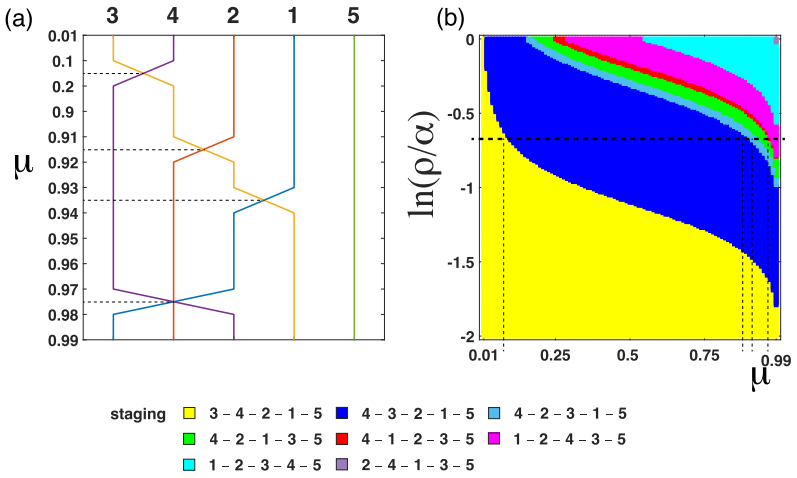

FIG. 10. (a) *A braid diagram generated using a graph in Figure* 1(a) *where* $\ln(\rho/\alpha) = -0.7$. *Dashed horizontal lines highlight examples of a change in staging from one threshold to the next.* (b) *A braid surface showing staging dynamics as parameter* ρ/α *varies (with a unique seed* $p_3(0) = \beta = 1/100$*).*

exact Braid surface and the Braid surfaces constructed using the three arrival time estimates. As in Figure 8b, the nonlinear approximation (Figure 11d) has the lowest distribution of error, followed by the Lambert distance (Figure 11c) and the linear estimate (Figure 11b). The errors are generally the most pronounced in the diffusion-dominated region, where $\log(\rho/\alpha) \approx 0$, and the accuracy of all methods tends to improve with a decrease in the threshold μ. The methods for arrival times have been designed under the general assumptions that the threshold is small and the dynamics is growth-dominated. This corresponds to the lower left-hand corned of the braid surface diagram. In that regime, we see that all methods are excellent in recovering the exact staging and can therefore be used systematically for that purpose. These general observations again hold true for the high resolution $N = 1015$ connectome (not shown).

6. Conclusion. Dynamical processes on a network bring together the combined effects of local dynamics, transport, global topology, and interactions between nodes to create complex patterns of evolution. Yet, in the case of a single autocatalytic growth process transported by the graph Laplacian, a simple picture emerges. For large ratio of the diffusion by the growth constant, all nodes in the graph are quickly seeded and the concentrations at all nodes behave in unison. However, when diffusion is small enough, compared to the autocatalytic growth, the network is systematically invaded, node by node, and the overall process can be characterized by an early seeding and by the arrival time, at each node, of a nonzero concentration.

In this manuscript, we have presented three different and complementary approaches to the problem that provide most of the relevant information about the dynamics. First, we discussed the linear solutions, obtained by a suitable linearization of the entire system of differential equations. We showed that these solutions provide a systematic lower bound for the arrival times. While these estimates have a systematic bias, the method is universal and can be easily generalized to other systems. The linear method also has the advantage of capturing the topology of the entire network through an exponential of the graph Laplacian. The error analysis provided in Appendix A shows that the method is robust and the error generated is independent of the choice of parameters, edge density, network size, initial condition,

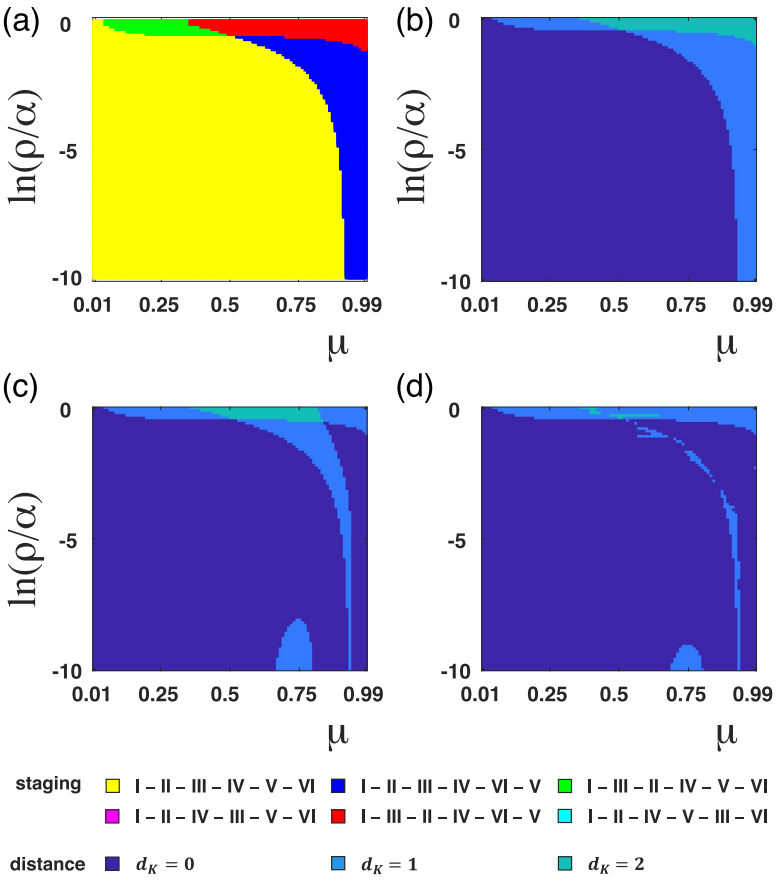

FIG. 11. (a) *An exact braid surface computed using* (1) *on the* $N = 83$ *brain connectome. Kendall tau errors for the approximated braid surfaces using* (b) *the linear method,* (c) *the Lambert method, and* (d) *the nonlinear method. An initial seeding is chosen so that the average concentration in the entorhinal cortex is less than* $1/100$*:* $p_{27} = p_{68} = 1/200$ *and* $p_i = 0$ *for all other* $i \in 1, \ldots, N$*. Yellow regions correspond to a canonical connectome Braak staging. (Color figure available online.)*

or topology. The primary drawback of the linear approach is that it requires the numerical solution of a transcendental equation at each node, which may be difficult in the absence of an initial guess.

Second, we introduced the notion of Lambert distance on a graph based on a linear estimate of transport between two neighboring nodes. From the Lambert distance, the arrival times can be estimated by finding the shortest path with respect to this metric. This method is particularly powerful as it is relies on standard algorithms implemented in most scientific computing frameworks in a robust and efficient way, as exemplified by the two-line code shown in Appendix C. It can be used to provide approximations of the dynamics. The Lambert method also leads to characteristic time scales, such as the denial and pandemic periods, and endows the network with a metric that is closely associated with the evolving dynamics of interest. The error analysis provided in Appendix A shows that the Lambert method is not sensitive to the choice of initial conditions, parameters, network size, and edge density, but is affected by the topology, especially in unweighted networks with multiple shortest paths between pairs of nodes.

Third, we developed an asymptotic method based on the nonlinear variation of constants. The asymptotic approach builds an approximate solution of the problem that is valid uniformly and has the correct asymptotic properties. This method further improves the Lambert estimates. The primary drawback of this method is that it relies on the existence of an exact solution for the single-node problem. As a result, this approach may fail to generalize to other problems.

Finally, the dynamics and the methods developed herein are motivated by the use of network models in the study of neurodegenerative diseases. Structural connectomes are often characterized by a small-world type topology with highly connected subnetworks and small mean-shortest path length. For the connectomes discussed here, it takes on average fewer than three edges to connect any two nodes. In this context, we have shown that analytical methods for computing the arrival times are extremely powerful and can capture essential features, such as the correct arrival times and Braak staging. We also shown that we can use the analytical solutions provided by these asymptotic expansions to compare the solutions of the model with clinical data on toxic tau concentrations. This analysis allowed us to identify typical times for both mild and advanced stages of the disease. It is therefore a flexible tool for investigation that offers more insight and flexibility than full computational studies while still including fundamental mechanisms.

Appendix A. Error analysis. To check the validity of our approach, we consider a number of parametric studies where we systematically vary one parameter and study the resulting error by comparing the exact numerical values with both Lambert and linear estimates. Since the asymptotic solution is an improvement over the Lambert estimates, we do not use it for this error analysis. We recall that given exact and approximate arrival times, we defined the error as

$$\mathcal{E}_{\text{approx}} = \frac{\sum_{i=1}^{N} |\widehat{\tau}_i - \tau_i|}{\sum_{i=1}^{N} \tau_i}, \qquad \text{approx} \in \{\text{linear, Lambert}\}. \tag{47}$$

For the rest of this section, we fix the growth constant $\alpha = 1/2$ and systematically vary the diffusion constant ρ, the initial condition, the weight of the edge connection, and the size of the network, and choose a few different one-parameter networks describing different topology.

A.1. Varying the diffusion constant. Since only the ratio ρ/α is meaningful, we fix $\alpha = 1/2$ and vary the constant ρ for the $N = 83$ network with diffusive weights and initial conditions: $p_{27} = p_{68} = 1/20$. We observe in Figure 12a that the error remains small and does not strongly depend on the value of ρ even when taken extreme. As expected when ρ is of order one, the Lambert estimates break down since we are not in the growth-dominated regime anymore and the diffusion-dominated case of section 2.2 becomes relevant for the computation of the arrival times.

A.2. Varying the weights. To understand the possible dependence of our method on the choice of edge weights, we generalize (39) by considering the one-parameter family of adjacency matrices for the $N = 83$ connectome:

$$A_{ij} = \frac{n_{ij}}{l_{ij}^{w}}, \quad i, j = 1, \ldots, N, \tag{48}$$

where we take $w \in [0, 2]$. These matrices cover all the choices previously made in the literature for the propagation of toxic proteins on the structural connectome ($w \in \{0, 1, 2\}$). We see in Figure 12b that the errors from both the Lambert and the linear methods remain small and do not depend on the choice of weights.

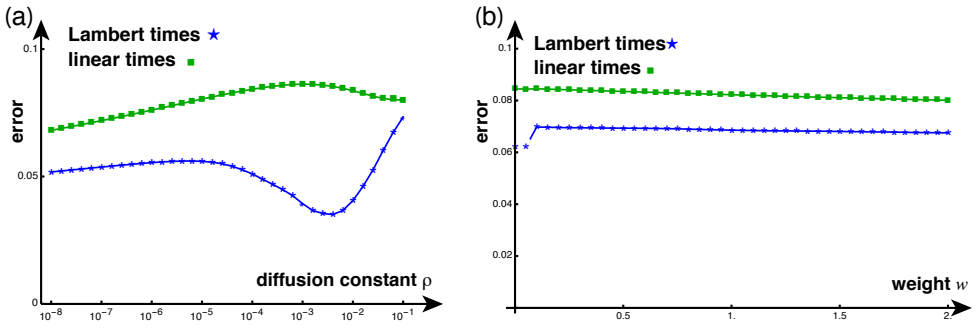

FIG. 12. *Error analysis with respect to* (a) *changes in weight* ($\rho = 0.01$/*year*) *and* (b) *changes in diffusion* ($w = 2$) *(parameters for both:* $\alpha = 0.5$/*year and initial conditions:* $p_{27}(0) = 1/20 = p_6 8(0)$ *and* $p_i = 0$ *for all other* $i \in \{1, \ldots, N\}$*). We observe that the method does not depend strongly on either the weight or the choice of parameters (as long as the system remains in the growth-dominated regime).*

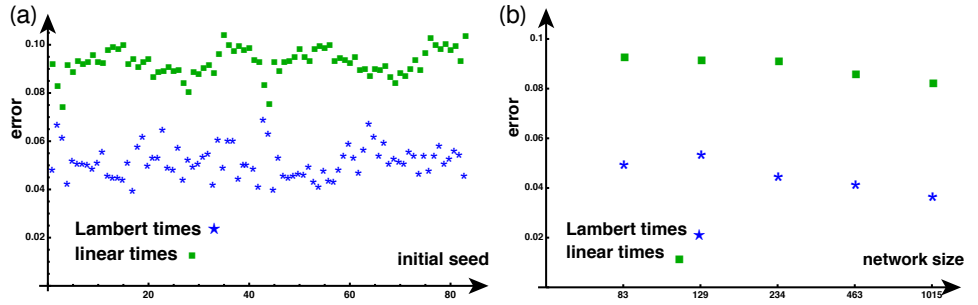

FIG. 13. *Error analysis with respect to changes in* (a) *initial conditions and* (b) *network sizes. Parameters for both:* $\alpha = 0.5$/*year,* $\rho = 0.01$/*year.* (a) *Initial conditions:* $p_i^{(i)} = 1/10$ *and* $p_j^{(i)} = 0$ *for all other* $j \in \{1, \ldots, 83\}$*.* (b) $N \in \{83, 129, 234, 463, 1015\}$ *and initial conditions:* $p_1 = 1/10$ *and* $p_j = 0$ *for* $j = 2, \ldots, N$*.*

A.3. Varying the initial conditions. To understand the possible dependence of our method on the choice of initial conditions, we consider the base case of the structural network with diffusive weight and $N = 83$ and systematically compute the error for different initial conditions by taking a nonzero concentration at a single-node i and setting all other initial conditions to zero. The result of the 83 runs is shown in Figure 13a. The error analysis shows that there is little effect associated with choices in initial conditions and that the typical error associated with the Lambert method is about 1/2 to 3/4 the error associated with the linear method.

A.4. Varying the network size. To understand the possible dependence of our method on the network size, we consider the five different resolutions of networks with diffusive weight and $N = 83, 129, 234, 463, 1015$ and systematically compute the error for each case. The results of the five runs are shown in Figure 13b. The error analysis shows that there is little effect associated with changes in network size.

A.5. Varying the topology. To understand the possible dependence of our method on the topology of the network, we consider different classes of networks used for this type of analysis. We consider two types of graphs: (a) a distribution of scale-free networks where a graph is built by adding a new vertex with k edges at each

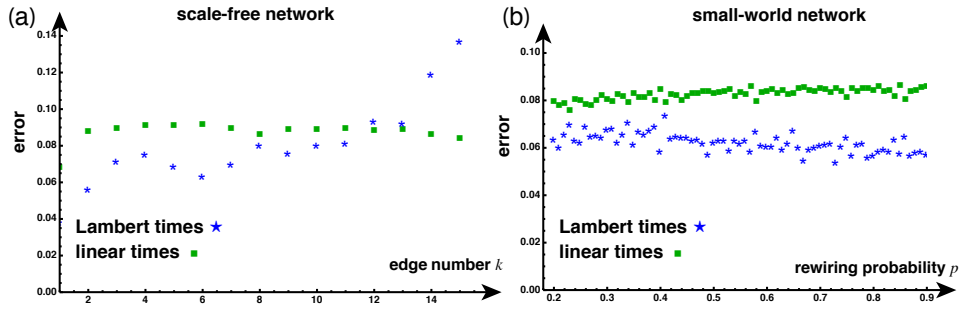

FIG. 14. *Error analysis with respect to changes in topology.* (a) *Scale-free network where* k *edges are added at each step.* (b) *Small-world network (with average degree of node* 10*) with rewiring probability* p*. Parameters:* $N = 500$*,* $\alpha = 1/\text{year}$*,* $\rho = 0.001/\text{year}$*. Initial conditions:* $p_1 = 1/10$ *and* $p_j = 0$ *for* $j = 2, \ldots, N$*.*

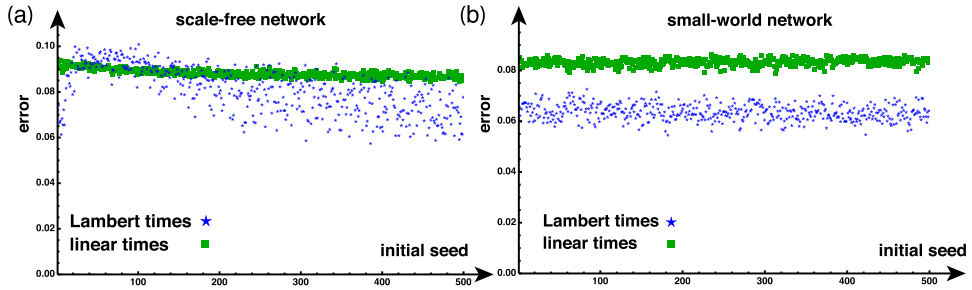

FIG. 15. *Error analysis with respect to changes in initial seeding for* (a) *a scale-free network where* 6 *edges are added at each step and* (b) *a small-world network (with average degree of node* 10*) with rewiring probability* 0.5*. Parameters:* $N = 500$*,* $\alpha = 1/\text{year}$*,* $\rho = 0.001/\text{year}$*. Initial conditions: unique seeding at* $p_j = 1/10$ *for* $j = 1, \ldots, N$*.*

step; and (b) a distribution of small-world networks with rewiring probability p. As shown in Figure 14, we observe that for scale-free networks, both the linear and the Lambert methods provide good estimates and that as the number of edges per vertex increases, the Lambert method is not as reliable due to the fact that multiple shortest paths are possible and the Lambert method only considers a single one of these. We also note in all the error analysis that the linear method is robust and that the error it generates does not depend on the topology.

A.6. Varying the initial seeding for different topology. To further study how the topology may influence the error for different seeding locations, we consider two different topologies, a scale-free network with edge number 6 and a small-world network with rewiring probability 0.5, and compute the error for each of the $N = 500$ possible single seed locations in Figure 15. We see no strong dependence of the errors on the seeding site, but there is an overall structure with respect to the seed ordering in Figure 15a that would be interesting to analyze further.

Appendix B. Variability. It is of interest to see whether our methods can capture the variability among different individuals. We consider the 426 individual connectomes included in the Budapest data set (see main text for details) and extract for each one a diffusion-weighted symmetric adjacency matrix. Since most of these connectomes are disconnected, we superimpose to each a small perturbation (taken to be 1/100) given by the average connectome and normalize the resulting adjacency matrix so that its maximum element is one. Then, for each connectome, we compute both the exact (numerical) mean arrival time and its Lambert approximation for the

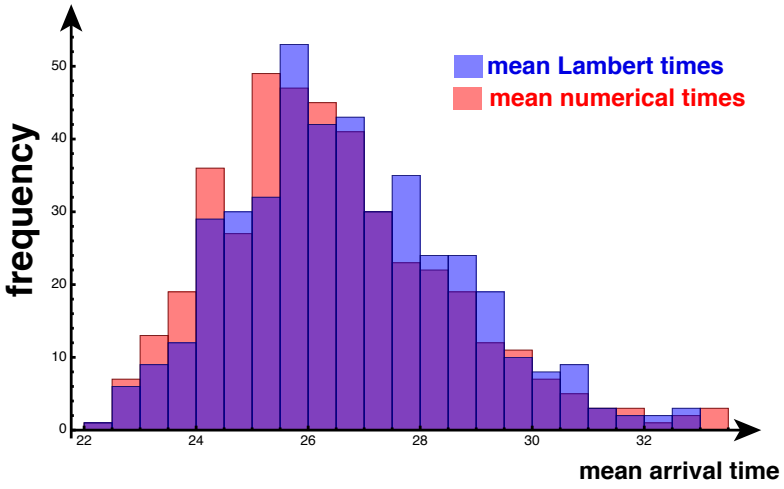

FIG. 16. *Histogram of the mean arrival times for each of the* 426 *structural connectomes in the reference data set obtained from either the exact (numerical) method or its Lambert approximation.*

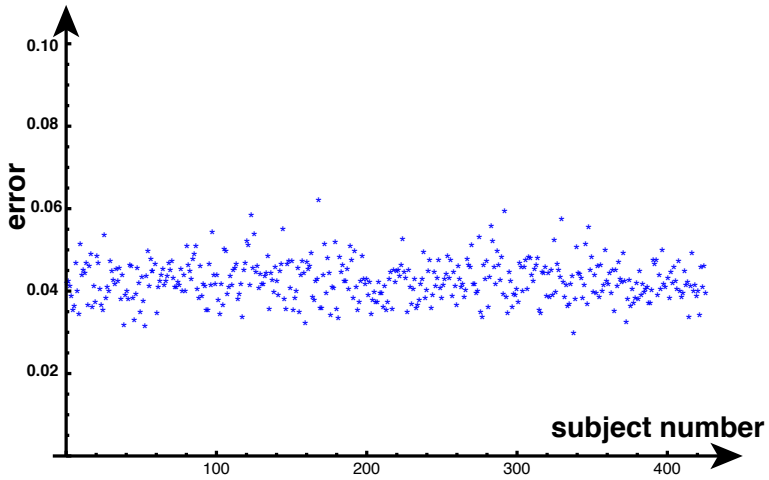

FIG. 17. *Error* $\mathcal{E}_{Lambert}$ *for each of the* 426 *subjects.*

same values of the parameters and initial conditions used in the main text ($N = 83, \alpha = 0.5$/year, $\rho = 0.01$/year. $p_{27} = p_{68} = 1/20$ and $p_i = 0$ for all other $i \in \{1, \ldots, N\}$). The distribution of arrival times is given in Figure 16 with approximated mean arrival time equal to 26.7 years (compared to 26.4 years for the exact method) and a standard deviation of 2.0 years for both methods. We conclude that the Lambert method captures both the mean and variability found in individual brain connectomes.

In order to see whether individual variability is captured, we also give, in Figure 17, the error for each subject. We see that the error remains below 0.065 and is of the same magnitude (between 3% and 6%) as in the other cases reported in sections A.1 and A.2, which suggests that the approximation is performing the same across individuals.

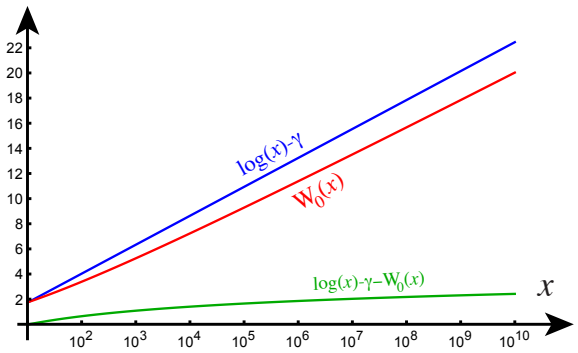

FIG. 18. *Comparison of the effective edge time (blue) and the Lambert edge time (red) as a function of their argument in the growth-dominated case. We observe that the use of the logarithm function for the distance creates a systematic overestimate of the arrival time of order one in the case of two nodes. In a network, this difference (shown in green) adds to every single edge along a path and causes a systematic error.*

Appendix C. Code. To appreciate how easy it is to implement the Lambert times for all single seeds, we give explicitly the entire self-explanatory code in Wolfram Mathematica 13.0, for a given adjacency matrix **A** given by the file `adjacency.csv`. On a standard laptop, it runs in about 0.1 second for $N = 83$ and in under 5 seconds for $N = 1015$:

```
A = Import[''adjacency.csv''];alpha=1/2;rho=1/10;beta=1/10;mu=10;
LambertTimes = Log[(1-beta)mu/(beta(1-mu))]/alpha+
  GraphDistanceMatrix[WeightedAdjacencyGraph[
                                    ProductLog[alpha/(rho A)]/alpha]]
```

The Lambert times for a single nonvanishing initial condition at position `seed` is then the corresponding column in the matrix: `LambertTimes[[seed]]`.

Appendix D. The effective distance and the Lambert distance. The idea to use the Lambert function as an approximation for the arrival times was first proposed in the seminal paper [12] in conjunction with an effective edge distance replacing (19) by

$$t_{\text{effective}} = \frac{1}{\alpha}\left(\log\left(\frac{\alpha}{\rho A_{ij}}\right) - \gamma\right), \tag{49}$$

where γ is the Euler constant. Surprisingly, despite showing better fit to the data (see Figure 2 in [12]), the Lambert distance was disregarded by the authors and the focus of the community moved to the notion of the effective distance based on the logarithmic edge time (49). Following the influential work of Brockman, Helbing and co-authors, who introduced the fundamental notion of a metric based on logarithmic edge times [13, 22], the effective distance gained further popularity, probably due to a familiarity with the logarithmic function. Despite some attempts to relate both notions [21], the two distances remain of order 1 even for very large values of the argument, as shown in Figure 18. The difference between the two estimates creates a systematic bias, and we found that the effective distance does not perform as well as the Lambert distance, especially on long paths.

We also note that the presence of the Euler constant in (49) is related to the stochastic approach of [12], which showed that, within the probabilistic framework, arrival times follow a Gumbel distribution. Within the context of a deterministic

model, however, there is no a priori justification for this constant, although it lingers in the literature like a vestigial appendage.

REFERENCES

[1] E. Kuhl, *Computational Epidemiology*, Springer, 2021.

[2] S. K. Kaufman, K. Del Tredici, T. L. Thomas, H. Braak, and M. I. Diamond, *Tau seeding activity begins in the transentorhinal/entorhinal regions and anticipates phospho-tau pathology in Alzheimer's disease and PART*, Acta Neuropathol., 136 (2018), pp. 57–67.

[3] S. Fornari, A. Schäfer, A. Goriely, and E. Kuhl, *Prion-like spreading of Alzheimer's disease within the brain's connectome*, J. Roy. Soc. Interface, 16 (2019), 20190356.

[4] S. Fornari, A. Schäfer, E. Kuhl, and A. Goriely, *Spatially-extended nucleation-aggregation-fragmentation models for the dynamics of prion-like neurodegenerative protein-spreading in the brain and its connectome*, J. Theor. Biol., 486 (2020), 110102.

[5] B. Goncalves, D. Balcan, and A. Vespignani, *Human mobility and the worldwide impact of intentional localized highly pathogenic virus release*, Sci. Rep., 3 (2013), pp. 1–7.

[6] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, *Epidemic processes in complex networks*, Rev. Modern Phys., 87 (2015), pp. 925–979.

[7] C. Hens, U. Harush, S. Haber, R. Cohen, and B. Barzel, *Spatiotemporal signal propagation in complex networks*, Nat. Phys., 15 (2019), pp. 403–412.

[8] A. Gautreau, A. Barrat, and M. Barthélemy, *Arrival time statistics in global disease spread*, J. Stat. Mech, 2007 (2007), L09001.

[9] K. Linka, P. Rahman, A. Goriely, and E. Kuhl, *Is it safe to lift COVID-19 travel bans? The Newfoundland story*, Comput. Mech., 66 (2020), pp. 1081–1092.

[10] A. Raj, A. Kuceyeski, and M. Weiner, *A network diffusion model of disease progression in dementia*, Neuron, 73 (2012), pp. 1204–1215.

[11] R. Burioni, S. Chibbaro, D. Vergni, and A. Vulpiani, *Reaction spreading on graphs*, Phys. Rev. E, 86 (2012), 055101.

[12] A. Gautreau, A. Barrat, and M. Barthelemy, *Global disease spread: Statistics and estimation of arrival times*, J. Theor. Biol., 251 (2008), pp. 509–522.

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

[14] A. Koher, H. H. K. Lentz, J. P. Gleeson, and P. Hövel, *Contact-based model for epidemic spreading on temporal networks*, Phys. Rev. X, 9 (2019), 031017.

[15] M. Schröder, X. Zhang, J. Wolter, and M. Timme, *Dynamic perturbation spreading in networks*, IEEE Trans. Netw. Sci. Eng., 7 (2019), pp. 1019–1026.

[16] X. Zhang, D. Witthaut, T. and M. Timme, *Topological determinants of perturbation spreading in networks*, Phys. Rev. Lett., 125 (2020), 218301.

[17] B. Zinner, G. A. Harris, and W. Hudson, *Traveling wavefronts for the discrete Fisher's equation*, J. Differential Equations, 105 (1993), pp. 46–62.

[18] A. Hoffman and M. Holzer, *Invasion fronts on graphs: The Fisher-KPP equation on homogeneous trees and Erdős-Réyni graphs*, Discrete Contin. Dyn. Syst. Ser. B, 24 (2019), pp. 671–694.

[19] G. Faye, *Traveling fronts for lattice neural field equations*, Phys. D, 378–379 (2018), pp. 20–32.

[20] L. M. Chen, M. Holzer, and A. Shapiro, *Estimating epidemic arrival times using linear spreading theory*, Chaos, 28 (2018), 013105.

[21] A. Jamieson-Lane and B. Blasius, *Calculation of epidemic arrival time distributions using branching processes*, Phys. Rev. E, 102 (2020), 042301.

[22] D. Brockmann and D. Helbing, *The hidden geometry of complex, network-driven contagion phenomena*, Science, 342 (2013), pp. 1337–1342.

[23] J. Weickenmeier, E. Kuhl, and A. Goriely, *Multiphysics of prionlike diseases: Progression and atrophy*, Phys. Rev. Lett., 121 (2018), 158101.

[24] R. Merris, *Laplacian matrices of graphs: A survey*, Linear Algebra Appl., 197 (1994), pp. 143–176.

[25] F. R. K. Chung, *Spectral Graph Theory*, CBMS Regional Conf. Ser. in Math. 92, American Mathematical Society, 1997.

[26] P. Putra, T. B. Thompson, P. Chaggar, and A. Goriely, *Braiding Braak and Braak: Staging patterns and model selection in network neurodegeneration*, Netw. Neurosci., 5 (2021), pp. 1–41.

[27] M. Asllani, B. R. Da Cunha, E. Estrada, and J. P. Gleeson, *Dynamics impose limits to detectability of network structure*, New J. Phys., 22 (2020), 063037.
[28] R. E. O'Malley, Jr., and E. Kirkinis, *Variation of parameters and the renormalization group method*, Stud. Appl. Math., 134 (2015), pp. 215–232.
[29] J.-L. Lagrange, *Mécanique Analytique*, Albert Blanchard, Paris, 1788.
[30] M. Jucker and L. C. Walker, *Propagation and spread of pathogenic protein assemblies in neurodegenerative diseases*, Nat. Neurosci., 21 (2018), pp. 1341–1349.
[31] T. T. Olsson, O. Klementieva, and G. K. Gouras, *Prion-like seeding and nucleation of intracellular amyloid-β*, Neurobiol. Dis., 113 (2018), pp. 1–10.
[32] M. Jucker and L. C. Walker, *Self-propagation of pathogenic protein aggregates in neurodegenerative diseases*, Nature, 501 (2013), pp. 45–51.
[33] F. Clavaguera, I. Lavenir, B. Falcon, S. Frank, M. Goedert, and M. Tolnay, *"Prion-like" templated misfolding in tauopathies*, Brain Pathol., 23 (2013), pp. 342–349.
[34] M. Goedert, *Alzheimer's and Parkinson's diseases: The prion concept in relation to assembled Aβ, tau, and α-synuclein*, Science, 349 (2015), 1255555.
[35] A. Mudher, M. Colin, S. Dujardin, M. Medina, I. Dewachter, S. M. A. Naini, E.-M. Mandelkow, E. Mandelkow, L. Buée, M. Goedert, and J.-P. Brion, *What is the evidence that tau pathology spreads through prion-like propagation?*, Acta Neuropathol. Commun., 5 (2017), 99.
[36] S. L. DeVos, B. T. Corjuc, D. H. Oakley, C. K. Nobuhara, R. N. Bannon, A. Chase, C. Commins, J. A. Gonzalez, P. M. Dooley, M. P. Frosch, and B. T. Hyman, *Synaptic tau seeding precedes tau pathology in human Alzheimer's disease brain*, Front. Neurosci., 12 (2018), 267.
[37] A. Raj, E. LoCastro, A. Kuceyeski, D. Tosun, N. Relkin, and M. Weiner; Alzheimer's Disease Neuroimaging Initiative (ADNI), *Network diffusion model of progression predicts longitudinal patterns of atrophy and metabolism in Alzheimer's disease*, Cell Rep., 10 (2015), pp. 359–369.
[38] S. Pandya, A. Kuceyeski, and A. Raj, *The brain's structural connectome mediates the relationship between regional neuroimaging biomarkers in Alzheimer's disease*, J. Alzheimers Dis., 55 (2017), pp. 1639–1657.
[39] S. Pandya, C. Mezias, and A. Raj, *Predictive model of spread of progressive supranuclear palsy using directional network diffusion*, Front. Neurol., 8 (2017), 692.
[40] S. Pandya, Y. Zeighami, B. Freeze, M. Dadar, D. L. Collins, A. Dagher, and A. Raj, *Predictive model of spread of Parkinson's pathology using network diffusion*, NeuroImage, 192 (2019), pp. 178–194.
[41] Y. Iturria-Medina, R. C. Sotero, P. J. Toussaint, A. C. Evans, and Alzheimer's Disease Neuroimaging Initiative, *Epidemic spreading model to characterize misfolded proteins propagation in aging and associated neurodegenerative disorders*, PLoS Comput. Biol., 10 (2014), e1003956.
[42] Y. Iturria-Medina, F. M. Carbonell, R. C. Sotero, F. Chouinard-Decorte, A. C. Evans, and Alzheimer's Disease Neuroimaging Initiative, *Multifactorial causal model of brain (dis)organization and therapeutic intervention: Application to Alzheimer's disease*, Neuroimage, 152 (2017), pp. 60–77.
[43] F. Carbonell, Y. Iturria-Medina, and A. C. Evans, *Mathematical modeling of protein misfolding mechanisms in neurological diseases: A historical overview*, Front. Neurol., 9 (2018), 37.
[44] J. W. Vogel, Y. Iturria-Medina, O. T. Strandberg, R. Smith, E. Levitis, A. C. Evans, and O. Hansson, *Spread of pathological tau proteins through communicating neurons in human Alzheimer's disease*, Nat. Commun., 11 (2020), pp. 1–15.
[45] Y.-Q. Zheng, Y. Zhang, Y. Yau, Y. Zeighami, K. Larcher, B. Misic, and A. Dagher, *Local vulnerability and global connectivity jointly shape neurodegenerative disease propagation*, PLoS Biol., 17 (2019), e3000495.
[46] M. Von Smoluchowski, *Drei Vorträge über Diffusion, Brownsche Bewegung und Koagulation von Kolloidteilchen*, Z. Phys., 17 (1916), pp. 557–585.
[47] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, *Nucleated polymerization with secondary pathways.* III. *equilibrium behavior and oligomer populations*, J. Chem. Phys., 135 (2011), 065107.
[48] S. I. A. Cohen, M. Vendruscolo, C. M. Dobson, and T. P. J. Knowles, *Nucleated polymerization with secondary pathways.* II. *Determination of self-consistent solutions to growth processes described by non-linear master equations*, J. Chem. Phys., 135 (2011), 065106.
[49] S. I. A. Cohen, M. Vendruscolo, M. E. Welland, C. M. Dobson, E. M. Terentjev, and T. P. J. Knowles, *Nucleated polymerization with secondary pathways.* I. *Time evolution of the principal moments*, J. Chem. Phys., 135 (2011), 065105

[50] G. Meisl, X. Yang, E. Hellstrand, B. Frohm, J. B. Kirkegaard, S. I. A. Cohen, C. M. Dobson, S. Linse, and T. P. J. Knowles, *Differences in nucleation behavior underlie the contrasting aggregation kinetics of the* Aβ40 *and* Aβ42 *peptides*, Proc. Natl. Acad. Sci., 111 (2014), pp. 9384–9389.

[51] M. Bertsch, B. Franchi, N. Marcello, M. C. Tesi, and A. Tosin, *Alzheimer's disease: A mathematical model for onset and progression*, Math. Med. Biol., 34 (2017), pp. 193–214.

[52] T. B. Thompson, G. Meisl, T. P. J. Knowles, and A. Goriely, *The role of clearance mechanisms in the kinetics of toxic pathological protein aggregation involved in neurodegenerative diseases*, J. Chem. Phys., 154 (2021), 125101.

[53] S. Fornari, A. Schäfer, M. Jucker, A. Goriely, and E. Kuhl, *Prion-like spreading of Alzheimer's disease within the brain's connectome*, J. Roy. Soc. Interface, 16 (2019), 20190356.

[54] A. Schäfer, P. Chaggar, A. Goriely, and E. Kuhl, *Correlating tau pathology to brain atrophy using a physics-based Bayesian model*, Eng. Comput., 38 (2022), pp. 3867–3877.

[55] A. Schäfer, P. Chaggar, T. B. Thompson, A. Goriely, E. Kuhl, and ADNI, *Predicting brain atrophy from tau pathology: A summary of clinical findings and their translation into personalized models*, Brain Multiphysics, 2 (2021), 100039.

[56] A. Whittington, D. J. Sharp, and R. N. Gunn, *Spatiotemporal distribution of β-amyloid in Alzheimer disease is the result of heterogeneous regional carrying capacities*, J. Nucl. Med., 59 (2018), pp. 822–827.

[57] J. A. McNab, B. L. Edlow, T. Witzel, S. Y. Huang, H. Bhat, K. Heberlein, T. Feiweier, K. Liu, B. Keil, J. Cohen-Adad, et al., *The human connectome project and beyond: Initial applications of* 300 *mT/m gradients*, NeuroImage, 80 (2013), pp. 234–245.

[58] B. Szalkai, C. Kerepesi, B. Varga, and V. Grolmusz, *Parameterizable consensus connectomes from the Human Connectome Project: The Budapest Reference Connectome Server v*3.0, Cogn. Neurodyn., 11 (2017), pp. 113–116.

[59] C. Kerepesi, B. Szalkai, B. Varga, and V. Grolmusz, *The braingraph.org database of high resolution structural connectomes and the brain graph tools*, Cogn. Neurodyn., 11 (2017), pp. 483–486.

[60] A. Daducci, S. Gerhard, A. Griffa, A. Lemkaddem, L. Cammoun, X. Gigandet, R. Meuli, P. Hagmann, and J.-P. Thiran, *The Connectome Mapper: An open-source processing pipeline to map connectomes with MRI*, PLoS One, 7 (2012), e48121.

[61] T. B. Thompson, P. Chaggar, E. Kuhl, and A. Goriely, for the Alzheimer's Disease Neuroimaging Initiative, *Protein-protein interactions in neurodegenerative diseases: A conspiracy theory*, PLoS Comput. Biol., 16 (2020), e1008267.

[62] A. Goriely, E. Kuhl, and C. Bick, *Neuronal oscillations on evolving networks: Dynamics, damage, degradation, decline, dementia, and death*, Phys. Rev. Lett., 125 (2020), 128102.

[63] P. Chaggar, T. Thompson, and A. Goriely, *Mathematical Models of Proteopathy and Atrophy in the Human Brain*, private report, 2020.

[64] A. Schäfer, E. C. Mormino, and E. Kuhl, *Network diffusion modeling explains longitudinal tau pet data*, Front. Neurosci., 14 (2020), 1370.

[65] S. Fornari, A. Schäfer, E. Kuhl, and A. Goriely, *Spatially-extended nucleation-aggregation-fragmentation models for the dynamics of prion-like neurodegenerative protein-spreading in the brain and its connectome*, J. Theor. Biol., 486 (2020), 110102.

[66] J. W. Vogel, A. L. Young, N. P. Oxtoby, R. Smith, et al., *Four distinct trajectories of tau deposition identified in Alzheimer's disease*, Nat. Med., 27 (2021), pp. 871–881.

[67] H. Braak and E. Braak, *Neuropathological stageing of Alzheimer-related changes*, Acta Neuropathol., 82 (1991), pp. 239–259.

[68] H. Braak, I. Alafuzoff, T. Arzberger, H. Kretzschmar, and K. Del Tredici, *Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry*, Acta Neuropathol., 112 (2006), pp. 389–404.

[69] A. Delacourte, J.-P. David, N. Sergeant, L. Buée, A. Wattez, P. Vermersch, F. Ghozali, C. Fallet-Bianco, F. Pasquier, F. Lebert, H. Petit, and C. Di Menza, *The biochemical pathway of neurofibrillary degeneration in aging and Alzheimer's disease*, Neurology, 52 (1999), pp. 1158–1165.

[70] H. Cho, J. Y. Choi, M. S. Hwang, Y. J. Kim, H. M. Lee, H. S. Lee, J. H. Lee, Y. H. Ryu, M. S. Lee, and C. H. Lyoo, *In vivo cortical spreading pattern of tau and amyloid in the Alzheimer disease spectrum*, Ann. Neurol., 80 (2016), pp. 247–258.

[71] Flortaucipir (AV-1451) processing methods, available at http://adni.loni.usc.edu/data-samples/access-data/ (accessed 20-12-2021).
[72] A. Raj, V. Tora, X. Gao, H. Cho, J. Y. Choi, Y. H. Ryu, C. H. Lyoo, and B. Franchi, *Combined model of aggregation and network diffusion recapitulates Alzheimer's regional tau-positron emission tomography*, Brain Connect., 11 (2021), pp. 624–638.
[73] G. R. Poudel, I. H. Harding, G. F. Egan, and N. Georgiou-Karistianis, *Network spread determines severity of degeneration and disconnection in Huntington's disease*, Hum. Brain Mapp., 40 (2019), pp. 4192–4201.
[74] D. Acosta, F. Powell, Y. Zhao, and A. Raj, *Regional vulnerability in Alzheimer's disease: The role of cell-autonomous and transneuronal processes*, Alzheimer's Dementia, 14 (2018), pp. 797–810.
[75] M. G. Kendall, *A new measure of rank correlation*, Biometrika, 30 (1938), pp. 81–93.