Front propagation and arrival times in networks with application to neurodegenerative diseases

Prama Putra, Hadrien Oliveri, Travis Thompson, and Alain Goriely
Mathematical Institute, University of Oxford

January 4, 2022

■ **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 question 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 system; the Lambert time, obtained by considering the interaction of two 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 insight and 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.

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, leading to a global pandemic for a disease, or to dementia for the brain? The *arrival-time problem* consists in determining the 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 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 take the view that the continuum system is a model of the dynamical process, independently of any underlying discrete process [15, 16]. The central question is then to obtain meaningful estimates about its dynamics 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 network 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. Then 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, linearize 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 easy 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, both in terms of implementation and in terms of 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 V representing regions of interest and edges $\mathcal{E}$ representing connections between these regions. If $p_i = p_i(t)$ 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

$$\frac{dp_i}{dt} = -\rho \sum_{j=1}^N L_{ij} p_j + \alpha p_i (1 - p_i), \quad i = 1, \dots, N, \quad (1)$$

$$p_i(0) = q_i, \quad i = 1, \dots, N, \quad (2)$$

where ρ, 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

$$L_{ij} = -A_{ij} + \delta_{ij} \sum_{k=1}^N A_{ik} \quad (3)$$

where δ_{ij} is the Kronecker delta ($\delta_{ij} = 1$ if $i = j$, 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 one the many normalized versions of the Laplacian which are often preferred in the mathematical study of graphs [25]. However, as discussed in [26], our choice of Laplacian is the only 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 $q_i = 0$, $i \neq s$. Then, a simple 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 $\tau_i \geq 0$ and finite with the equality $\tau_i = 0$ attained only if $i = s$ and $\mu = \beta$). Unless specifically stated, we will mostly study the case where $\mu = \beta$ and the problem is to know 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, let $\beta = \max_i \{q_i\} > 0$ and 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 non-vanishing: 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 ($\rho = 0$) follows locally a logistic equation, with a sigmoid

solution. Hence, for all values $0 < \beta < 1$ of the seed, the system will evolve towards the toxic state, which implies that the arrival times τ_i are all positive and finite.

By construction, the graph Laplacian has one zero eigenvalue. The dynamics is therefore mostly controlled by λ_2 , the smallest non-zero eigenvalue of $\mathbf{L}$. Indeed, we can define two main regimes of interest for this system as shown in Fig. 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 Fig. 1 with weighted adjacency matrix

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

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

Figure 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 feeling 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: $\alpha = 1, \rho = 1/100$).

Intuitively, one expects that in the diffusion-dominated regime, shown in Fig. 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 Fig. 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 the arrival times as 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 and integrate it. In this case, the system reduces to

$$\frac{dp_i}{dt} = -\rho \sum_{j=1}^N L_{ij} p_j + \alpha p_i, \quad i = 1, \dots, N, \quad (5)$$

$$p_i(0) = q_i, \quad i = 1, \dots, N. \quad (6)$$

A compact solution to this problem is obtained in terms of matrix exponentials. Let $\mathbf{p}(t) = (p_1, \dots, p_N)$, $\mathbf{q} = \mathbf{p}(0) = (q_1, \dots, q_N)$, and $\mathbf{M} = -\rho \mathbf{L} + \alpha \mathbf{1}$, where $\mathbf{1}$ denotes the identity matrix. Then, the solution is given by

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

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) eigenvalues λ_i , $i \in \{1, \dots, N\}$ so that

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

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

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

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

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

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

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

A comparison of the linear solutions with the numerical solutions for our 5-node network is given in Fig. 2bc. 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 $\mu \rightarrow 0$ with $\mu > 0$, this approximation converges to the exact value as the nonlinear terms becomes negligible. This behavior is illustrated Fig. 2 where we compute the error made as a function of the seed size:

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

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

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 equation 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 the next section.

Fourth, this method is often overlooked at the 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 constitutes a robust universal method valid for small seed concentrations to obtain arrival times as further demonstrated in the next sections.

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

$$\frac{dp}{dt} = \alpha p(1 - p), \quad p(0) = \mu/N. \quad (13)$$

The solution of this equation is

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

with an arrival time at all nodes given by

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

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

Figure 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 grey rectangle, for the arrival time as well (b).

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

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

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.

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

An second approach is to define a notion of distance between nodes 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 two-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 $\mathbf{L}_{ij}$ 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

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

Expanding for small ρ , we obtain

$$\rho A_{ij} t e^{\alpha t} = 1. \quad (18)$$

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 $t = W_0(z)$):

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

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 $\Gamma_{ij} = (\gamma_0, \gamma_1, \dots, \gamma_n)$ with $\gamma_0 = i$ and $\gamma_n = j$ be this shortest path, then

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

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

$$\frac{dp}{dt} = \alpha p(1 - p), \quad p(0) = \beta,$$

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

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

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

$$\tilde{\tau}_{ij} = t_{ii} + W_{ij}. \quad (22)$$

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 $\mu = \beta$ a priori. 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 to be 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 reduces 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. Hence, 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

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

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

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

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 Fig. 1. The matrices corresponding to the Lambert edge distance and Lambert distance, respectively, are:

$$[t_{ij}] = \begin{bmatrix} \infty & 3.39 & 5.06 & \infty & \infty \\ 3.39 & \infty & 3.93 & \infty & \infty \\ 5.06 & 3.93 & \infty & 3.93 & 5.06 \\ \infty & \infty & 3.93 & \infty & 5.25 \\ \infty & \infty & 5.06 & 5.25 & \infty \end{bmatrix}, \quad W = \begin{bmatrix} 0.00 & 3.39 & 5.06 & 8.99 & 10.10 \\ 3.39 & 0.00 & 3.93 & 7.86 & 8.99 \\ 5.06 & 3.93 & 0.00 & 3.93 & 5.06 \\ 8.99 & 7.86 & 3.93 & 0.00 & 5.25 \\ 10.10 & 8.99 & 5.06 & 5.25 & 0.00 \end{bmatrix}, \quad (25)$$

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 Fig. 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.

Finally, we introduce the notion of *step distance*. The step distance is a vector, γ , whose j^{th} 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 Fig. 1, with seeding at node 1 we have $\gamma = (0, 1, 1, 2, 2)$. Thus, for example, this particular γ indicates that shortest path, with respect to the Lambert metric established by $\mathbf{W}$, between seed node 1 and target node 3 is comprised of 2 edges.

Figure 4: a. Exact (red), 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 $0 \leq \max_i \{q_i\} < \mu \ll 1$. 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

$$\tau_{ij} = \frac{1}{\alpha} \log \left(\frac{(1 - q_i)\mu}{(1 - \mu)q_i} \right) + W_{ij}, \quad (26)$$

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

$$\tilde{p}_j^{(i)}(t) = \mu \frac{t \exp(\alpha t)}{\tau_{ij} \exp(\alpha \tau_{ij})}. \quad (27)$$

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

$$\tilde{p}_j(t) = \sum_{i=1}^N \tilde{p}_j^{(i)}(t) = \mu t \exp(\alpha t) \sum_{i=1}^N \frac{1}{\tau_{ij}} \exp(-\alpha \tau_{ij}), \quad (28)$$

which reaches the level μ at time

$$\tilde{t}_j(\mu) = \frac{1}{\alpha} W_0 \left(\frac{\alpha}{\sum_{i=1}^N \frac{1}{\tau_{ij}} \exp(-\alpha \tau_{ij})} \right). \quad (29)$$

Just as in the case of a single seeding site, the estimate (29) can be used as an initial guess to compute the linear arrival times for the case of multiple seeding locations.

3.2 Discussion and comparison with other notions of 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

$$t_{\text{effective}} = \frac{1}{\alpha} \left(\log \left(\frac{\alpha}{\rho A_{ij}} \right) - \gamma \right), \quad (30)$$

where γ is the Euler constant. Surprisingly, despite showing better fit to the data (See Fig. 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 (30). 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], there is no meaningful limit in which the two distances coincide as can be observed directly from the graphs of these two function shown in Fig. 5. 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 (30) is related to the stochastic approach of [12] who 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 despite the fact that it lingers in the literature like a vestigial appendage.

Figure 5: 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.

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 method 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 solve 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

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

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 constants¹

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

where γ_i is the step distance between 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 γ .

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

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

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

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

that can be easily integrated to obtain

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

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

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

and the variables

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

¹The 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 [27], this method was first proposed by Lagrange in 1788 [28], almost exactly along the lines used here.

then, these new variables are solutions of the system

$$\frac{dZ_i}{dt} = \sum_{j=1}^N M_{ij} Z_j, \quad Z_i(0) = 0, \quad i \neq s. \quad (38)$$

The asymptotic solutions, for the 5-node network of Fig. 1, are shown in Fig. 6 together with the computations of the arrival times.

Figure 6: a. Exact (red), 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).

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. Indeed, it is now well appreciated that neurodegenerative diseases such as Parkinson's and Alzheimer's are associated with toxic proteins forming large aggregates. Starting in an initial region, these toxic proteins can systematically invade the brain through the axonal pathways known as 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 obtained from 426 patient data graphs part of the Budapest Reference Connectome v3.0 [29, 30] and generated using a multi-resolution Lausanne anatomical parcellation [31, 32]. The connectome networks are freely available [31] at five levels of resolution and we will use averaged connectome with the lowest ($N = 83$) and highest ($N = 1015$) resolution

[33]. These networks have edges with weight A_{kj} between nodes k and j , defined as the ratio of the number of axonal fibers between the nodes and the mean fiber length squared. 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 [34, 35, 33]. 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 to understand the disease as 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 post-mortem histopathology, suggesting that the progression is in the growth-dominated regime with a small diffusion constant [36, 37].

Therefore, we model the propagation of toxic tau protein through the network model of Fisher-KPP and used our methods to understand the dynamics.

5.1 Time scales of disease progression

In Fig. 7, 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*:

$$T_{\text{loc}} = 1/\alpha, \quad (39)$$

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

$$T_0 = T_{\text{loc}} \log \left(\frac{1 - \beta}{\beta} \right). \quad (40)$$

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

$$T_1(s) = \frac{1}{\alpha} W_0 \left(\frac{\alpha}{\rho a_s} \right), \quad a_s = \max_i A_{is}, \quad (41)$$

where a_s is the maximum of the s th 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

$$T_1 = \min_s T_1(s) = \frac{1}{\alpha} W_0 \left(\frac{\alpha}{\rho} \right), \quad (42)$$

as the *denial period*. The *denial period* (42) 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 [38, 34, 39, 40] and the potential alternative seeding sites referenced in connected to non-standard Alzheimer subtypes [41]. Conversely, the global denial period (42) represents an estimate of the minimal window of

Figure 7: Time scale estimates for the $N = 83$ connectome. Following the local growth at the seeding node with a time scale T_0 , the first neighboring is invaded at time T_1 . Successive nodes are invaded progressively with a general time scale depending on the number of steps to the last node. Parameter: $\alpha = 1/\text{year}$, $\rho = 0.001/\text{year}$. Initial condition: $p_{27} = 1/10$ and $p_i = 0$ for all other $i \in \{1, \dots, N\}$. 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).

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 Fig. 10. We define, for $i > 1$, $T_i(s) = iT_1(s)$ as 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 Fig. 10 characterize nodes whose shortest path to the seeding node has two edges. We see from Fig. 10 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:

$$T_{\text{max}} = (\ell + 1) T_1, \quad (43)$$

where ℓ is the maximal number of edges between any 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 $T_1(s)$. For instance, for the $N = 83$ connectome, we have $\ell = 3$, setting $\alpha = 1 \text{ year}^{-1}$, $s = 27$, and $T_{\text{max}} = 35$ years suggests taking the

diffusion constant to be

$$\rho = \frac{\exp\left(-\frac{\alpha T_{\max}}{\ell + 1}\right)}{T_{\max} \max_i A_{is}} \approx 0.0012/\text{year}. \quad (44)$$

5.2 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 post-mortem studies. In particular, histopathological studies have noted that neurofibrillary tangle follow a six-stage sequence [42, 43], called the *six Braak stages* [44]. While generalized progression sequences have also been proposed [45, 46], here we limit ourselves to the six Braak regions used in the standardized positron emission tomography AV-1451 data pipeline [47] of the Alzheimer's Disease Neuroimaging Initiative. In line with histopathological observation [42, 43] 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. Fig. 8 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.

Figure 8: Starting at the entorhinal cortex, the dynamics of the network Fisher-KPP equations recovers some of the key aspects of Braak staging $N = 83$, $\alpha = 1/\text{year}$, $\rho = 5.64 \times 10^{-3}/\text{year}$. Initial conditions: $p_{27} = p_{68} = 1/100$ and $p_i = 0$ for all other $i \in \{1, \dots, N\}$. Each of the six colors corresponds to the cortical regions that comprise a Braak stage. For instance, regions that are part of Braak I are in red.

We apply the three methods for arrival times on this problem for the same parameters as in Fig. 8 and compute the approximate linear solutions. In Fig. 9, 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 in divided by the number of nodes that region). As expected, the nonlinear solution provides a decent approximation of the solution and the best arrival times with comparative scores (12) given by

$$\mathcal{E}_{\text{lin}}(1/10) \approx 101, \quad \mathcal{E}_{\text{Lambert}}(1/10) \approx 41, \quad \mathcal{E}_{\text{nonlin}}(1/10) \approx 25. \quad (45)$$

Figure 9: 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 Fig. 8). Initial conditions 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 $i \in \{1, \dots, N\}$.

The same trends appears, again, in the highest resolution connectome, with 1015 nodes shown in Fig. 10, and the approximate nonlinear solutions provide a good, and best overall, match for both the arrival times and the concentration in the six Braak regions.

Figure 10: Arrival times at $\mu = 1/10$ for $N = 1015$ connectome (left) and average concentration (right) in each Braak region (parameters and color coding coincide with Fig. 8). Initial conditions chosen so that the total concentration in the entorhinal cortex is $1/10$: $p_{412} = p_{413} = p_{917} = p_{918} = p_{919} = 1/50$ and $p_i = 0$ for all other $i \in \{1, \dots, N\}$.

5.3 Braid diagrams and braid surfaces

We have applied each approximation method on the lowest and higher 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, Fig. 11(a) shows a simple example of a braid diagram for our exemplar 5-node network given in Fig.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 three is the first one at which the concentration reaches $1/100$ followed by node four, then two, one and five. For 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 associate a color label to that integer. For our example, the braid surface in Fig. 11(b) gives eight possible staging 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.

Figure 11: (a) A braid diagram generated using a graph in Fig. 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 $p_3(0) = \beta = 1/100$)

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 .

Figure 12: (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 $i \in 1, \dots, N$. Yellow regions correspond to a canonical connectome Braak staging.

To test the use of the Lambert distance to compute the braid surfaces, we use a the *Kendall tau* distance [48] to measure the distance between two stagings as the minimum number of transpositions necessary to transform one staging into the other. For instance, the Kendall tau distance between staging (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 non-linear estimation methods. Fig. 12(a) 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 [42, 43]. Figures 12(b)-(d) show the Kendall Tau

distance between the exact Braid surface and the Braid surfaces constructed using the three arrival time estimates. As in Fig. 8b, the non-linear approximation (Fig. 12(d)) has the lowest distribution of error, followed by the Lambert distance (Fig. 12(c)) and the linear estimate (Fig. 12(b)). 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 corner 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. 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 non-zero 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 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 easy to compute and can be used to provide overall approximation 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.

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 sub-networks and small mean-shortest path length. For the connectomes discussed here, it takes on average less than 3 edges to connect any two nodes in both the $N = 83$ and $N = 1015$ networks. In this context, we have shown that analytical methods to estimate and compute the arrival times are extremely powerful and can capture essential features, such as the correct arrival times and Braak staging.

Acknowledgment

The work of AG, HO and TT was supported by the Engineering and Physical Sciences Research Council grant EP/R020205/1 to AG. The work of TT was supported partially the John Fell Oxford University Press Research Fund grant 000872 (project code BKD00160) to TT.

References

- [1] Ellen Kuhl. *Computational Epidemiology*. Springer, 2021.
- [2] Sarah K Kaufman, Kelly Del Tredici, Talitha L Thomas, Heiko Braak, and Marc I Diamond. Tau seeding activity begins in the transentorhinal/entorhinal regions and anticipates phospho-tau pathology in alzheimer’s disease and part. *Acta neuropathologica*, 136(1):57–67, 2018.
- [3] Sveva Fornari, Amelie Schäfer, Alain Goriely, and Ellen Kuhl. Prion-like spreading of Alzheimer’s disease within the brain’s connectome. *Journal of The Royal Society Interface*, 16(159):20190356, 2019.
- [4] Sveva Fornari, Amelie Schäfer, Ellen Kuhl, and Alain Goriely. Spatially-extended nucleation-aggregation-fragmentation models for the dynamics of prion-like neurodegenerative protein-spreading in the brain and its connectome. *Journal of Theoretical Biology*, 486:110102, 2020.
- [5] Bruno Goncalves, Duygu Balcan, and Alessandro Vespignani. Human mobility and the worldwide impact of intentional localized highly pathogenic virus release. *Scientific Reports*, 3(1):1–7, 2013.
- [6] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.
- [7] Chittaranjan Hens, Uzi Harush, Simi Haber, Reuven Cohen, and Baruch Barzel. Spatiotemporal signal propagation in complex networks. *Nature Physics*, 15(4):403–412, 2019.
- [8] Aurélien Gautreau, Alain Barrat, and Marc Barthélemy. Arrival time statistics in global disease spread. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(09):L09001, 2007.
- [9] Kevin Linka, Proton Rahman, Alain Goriely, and Ellen Kuhl. Is it safe to lift COVID-19 travel bans? The Newfoundland story. *Computational Mechanics*, 66:1081–1092, 2020.
- [10] Ashish Raj, Amy Kuceyeski, and Michael Weiner. A network diffusion model of disease progression in dementia. *Neuron*, 73(6):1204–1215, 2012.
- [11] Raffaella Burioni, Sergio Chibbaro, Davide Vergni, and Angelo Vulpiani. Reaction spreading on graphs. *Physical Review E*, 86(5):055101, 2012.
- [12] Aurélien Gautreau, Alain Barrat, and Marc Barthelemy. Global disease spread: statistics and estimation of arrival times. *Journal of theoretical biology*, 251(3):509–522, 2008.
- [13] Flavio Iannelli, Andreas Koher, Dirk Brockmann, Philipp Hövel, and Igor M Sokolov. Effective distances for epidemics spreading on complex networks. *Physical Review E*, 95(1):012313, 2017.
- [14] Andreas Koher, Hartmut HK Lentz, James P Gleeson, and Philipp Hövel. Contact-based model for epidemic spreading on temporal networks. *Physical Review X*, 9(3):031017, 2019.

- [15] Malte Schröder, Xiaozhu Zhang, Justine Wolter, and Marc Timme. Dynamic perturbation spreading in networks. *IEEE Transactions on Network Science and Engineering*, 7(3):1019–1026, 2019.
- [16] Xiaozhu Zhang, Dirk Witthaut, Marc Timme, et al. Topological determinants of perturbation spreading in networks. *Physical Review Letters*, 125(21):218301, 2020.
- [17] Bertram Zinner, Gregory A. Harris, and William Hudson. Traveling Wavefronts for the Discrete Fisher’s Equation. *Journal of differential equations*, 105(1):46–62, 1993.
- [18] Aaron Hoffman and Matt Holzer. Invasion fronts on graphs: The Fisher-KPP equation on homogeneous trees and Erdős-Rényi graphs. *Discrete & Continuous Dynamical Systems-Series B*, 24(2):671–694, 2019.
- [19] Grégory Faye. Traveling fronts for lattice neural field equations. *Physica D: Nonlinear Phenomena*, 378-379:20–32, 2018.
- [20] Lawrence M Chen, Matt Holzer, and Anne Shapiro. Estimating epidemic arrival times using linear spreading theory. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(1):013105, 2018.
- [21] Alastair Jamieson-Lane and Bernd Blasius. Calculation of epidemic arrival time distributions using branching processes. *Physical Review E*, 102(4):042301, 2020.
- [22] Dirk Brockmann and Dirk Helbing. The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342(6164):1337–1342, 2013.
- [23] Johannes Weickenmeier, Ellen Kuhl, and Alain Goriely. Multiphysics of Prionlike Diseases: Progression and Atrophy. *Physical Review Letters*, 121(15):158101, 2018.
- [24] Russell Merris. Laplacian matrices of graphs: a survey. *Linear algebra and its applications*, 197:143–176, 1994.
- [25] Fan RK Chung and Fan Chung Graham. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.
- [26] Prama Putra, Travis B Thompson, Pavanjit Chaggar, and Alain Goriely. Braiding Braak and Braak: Staging patterns and model selection in network neurodegeneration. *Network Neuroscience*, 5(4):1–41, 2021.
- [27] Robert E O’Malley Jr and Eleftherios Kirkinis. Variation of parameters and the renormalization group method. *Studies in Applied Mathematics*, 134(2):215–232, 2015.
- [28] Joseph-Louis Lagrange. *Mécanique analytique*. Albert Blanchard, Paris, 1788.
- [29] Jennifer A McNab, Brian L Edlow, Thomas Witzel, Susie Y Huang, Himanshu Bhat, Keith Heberlein, Thorsten Feiweier, Kecheng Liu, Boris Keil, Julien Cohen-Adad, et al. The Human Connectome Project and beyond: initial applications of 300 mT/m gradients. *NeuroImage*, 80:234–245, 2013.
- [30] Balázs Szalkai, Csaba Kerepesi, Bálint Varga, and Vince Grolmusz. Parameterizable consensus connectomes from the human connectome project: The Budapest Reference Connectome Server v3. 0. *Cognitive Neurodynamics*, 11(1):113–116, 2017.

- [31] Csaba Kerepesi, Balázs Szalkai, Bálint Varga, and Vince Grolmusz. The braingraph.org database of high resolution structural connectomes and the brain graph tools. *Cognitive Neurodynamics*, 11:483–486, 2017.
- [32] Alessandro Daducci, Stephan Gerhard, Jean-Philippe Thiran, et al. The Connectome Mapper: An Open-Source Processing Pipeline to Map Connectomes with MRI. *PLoS One*, 7(12):e48121, 2012.
- [33] Travis B Thompson, Pavanjit Chaggar, Ellen Kuhl, Alain Goriely, Alzheimer’s Disease Neuroimaging Initiative, et al. Protein-protein interactions in neurodegenerative diseases: A conspiracy theory. *PLoS computational biology*, 16(10):e1008267, 2020.
- [34] Alain Goriely, Ellen Kuhl, and Christian Bick. Neuronal oscillations on evolving networks: dynamics, damage, degradation, decline, dementia, and death. *Physical Review Letters*, 125(12):128102, 2020.
- [35] Pavan Chaggar, Travis Thompson, and Alain Goriely. Mathematical models of proteopathy and atrophy in the human brain. *Private Report*, 2020.
- [36] Amelie Schäfer, Elizabeth C Mormino, and Ellen Kuhl. Network diffusion modeling explains longitudinal tau pet data. *Frontiers in Neuroscience*, 14:1370, 2020.
- [37] Amelie Schäfer, Pavanjit Chaggar, Travis B Thompson, Alain Goriely, Ellen Kuhl, Alzheimer’s Disease Neuroimaging Initiative, et al. Predicting brain atrophy from tau pathology: A summary of clinical findings and their translation into personalized models. *Brain Multiphysics*, 2:100039, 2021.
- [38] Sveva Fornari, Amelie Schäfer, Mathias Jucker, Alain Goriely, and Ellen Kuhl. Prion-like spreading of Alzheimer’s disease within the brain’s connectome. *Journal of the Royal Society Interface*, 16(159):20190356, 2019.
- [39] Travis B Thompson, Pavanjit Chaggar, Ellen Kuhl, Alain Goriely, and Alzheimer’s Disease Neuroimaging Initiative. Protein-protein interactions in neurodegenerative diseases: a conspiracy theory. *PLoS computational biology*, 16(10):e1008267, 2020.
- [40] Sveva Fornari, Amelie Schäfer, Ellen Kuhl, and Alain Goriely. Spatially-extended nucleation-aggregation-fragmentation models for the dynamics of prion-like neurodegenerative protein-spreading in the brain and its connectome. *Journal of Theoretical Biology*, 486:110102, 2020.
- [41] Jacob W. Vogel, Alexandra L. Young, Neil P. Oxtoby, Oskar Hannsson, et al. Four distinct trajectories of tau deposition identified in alzheimer’s disease. *Nature Medicine*, 27(5):871–881, 2021.
- [42] Heiko Braak and Eva Braak. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathologica*, 82(4):239–259, 1991.
- [43] Heiko Braak, Irina Alafuzoff, Thomas Arzberger, Hans Kretschmar, and Kelly Del Tredici. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathologica*, 112(4):389–404, 2006.
- [44] Sarah L DeVos, Bianca T Corjuc, Derek H Oakley, Chloe K Nobuhara, Riley N Bannon, Alison Chase, Caitlin Commins, Jose A Gonzalez, Patrick M Dooley, Matthew P Frosch, et al. Synaptic tau seeding precedes tau pathology in human alzheimer’s disease brain. *Frontiers in Neuroscience*, 12:267, 2018.

- [45] André Delacourte, Jean-Philippe David, Nicolas Sergeant, L Buee, A Wattez, P Vermer-sch, F Ghozali, C Fallet-Bianco, F Pasquier, F Lebert, et al. The biochemical pathway of neurofibrillary degeneration in aging and Alzheimer’s disease. *Neurology*, 52(6):1158–1158, 1999.
- [46] Hanna Cho, Jae Yong Choi, Mi Song Hwang, You Jin Kim, Hye Mi Lee, Hye Sun Lee, Jae Hoon Lee, Young Hoon Ryu, Myung Sik Lee, and Chul Hyoun Lyoo. In vivo cortical spreading pattern of tau and amyloid in the Alzheimer disease spectrum. *Annals of Neurology*, 80(2):247–258, 2016.
- [47] Flortaucipir (AV-1451) processing methods. <http://adni.loni.usc.edu/data-samples/access-data/>. Accessed: 20-12-2021.
- [48] M.G. Kendall. A New Measure of Rank Correlation. *Biometrika*, 30(1):81–93, 1938.