Tauopathy Staging Dynamics of Neurodegenerative Disease on Network-Organised Reaction-Diffusion

Prama Putra
Trinity College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2023

$\mathcal{QS}$ 96 : 1

Reading is the beginning, learning is a lifelong journey.

Acknowledgements

There are many people without whom this thesis would have never been possible. First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Alain Goriely, who has patiently guided me throughout the DPhil journey and has provided a near-endless stream of insight, knowledge, motivation, and wisdom. The lessons he taught me have been invaluable to me and I will always cherish the time he spent teaching me. Secondly, I would like to extend my gratitude to Dr. Travis Thompson and Dr. Hadrien Oliveri for their guidance and insightful discussions whom I admire as young scientists. Your dedication to unearthing the hidden knowledge has shown me the joy of learning. Thirdly, I am deeply grateful to The Jardine Foundation for supporting my DPhil study at the University of Oxford. Without their support, I would not have been able to broaden my horizons and explore the world.

On a personal note, Dr. Nuning Nuraini, Prof. Hadi Susanto, and Prof. Edy Soewono have been important teachers and mentors for my professional development. I would like to express my heartfelt gratitude for all the guidance and opportunities that you have provided me. I am deeply indebted to each and every one of you.

Any thesis acknowledgement section would be incomplete without mention of familial persuasion. Towards this end, I would like to thank my family, especially my mother who always supports me whenever and wherever. Your prayer has sent me to the door of the world. Finally, a sincere avowal of appreciation is owed to Rina Asri Aisyah, whom I will marry after finishing DPhil, for supporting me through the end of my DPhil journey and enlightening me on the meaning of having a DPhil for our own future family.

Abstract

Neurodegenerative diseases are a group of incurable disorders that are identified by the progressive loss of cognitive function and degeneration of the nervous system. Disorders such as Alzheimer's disease show a systematical progression of the disease's biological markers. A key finding in Alzheimer's disease is the hierarchical accumulation of toxic proteins which leads to a disease staging identification called staging. Many studies have used mathematical and computational models to explore the transport of toxic proteins. In this thesis, we investigate mathematical network models based on the prion-like hypothesis coupled with a linear transport principle to understand key features of the dynamics and extract information about toxic tau propagation. We consider three key problems. We use the Fisher-Kolmogorov-Petrovsky-Piskunov equation as a paradigm for the dynamics. First, we examine a model selection. We introduce a novel approach, based on braid diagrams, for studying the structured progression of tau protein evolving on a network. Second, we extract arrival time estimates for the dynamics. We present and compare three different estimates for the arrival time: (a) the linear arrival time obtained by linearizing the underlying system, (b) the Lambert time obtained by considering the interaction of two nodes, and (c) the nonlinear arrival time obtained by asymptotic techniques. Finally, we examine the role of brain regional heterogeneity and consider directed networks based on asymmetric brain connectivities. We study staging and arrival time for both heterogeneous and directed network systems. Overall, this study on tau protein propagation on networks with respect to staging and time of arrival provides new tools to investigate systematically the spatiotemporal evolution of toxic tau protein in the brain.

Contents

List of Figures

Chapter 1

Introduction

1.1 Neurodegenerative disease

It is estimated that around 50 million people are suffering from neurodegenerative diseases worldwide, with 850,000 people in the UK alone [177]. They carry a significant burden to the patient's families, carers, and society. Alzheimer's disease (AD) and Parkinson's disease (PD) are the most common neurodegenerative diseases, with AD cases estimated to be 60-70% of all cases [176].

Neurodegenerative diseases are a group of disorders that are primarily affecting the nervous system. Many diseases result in progressive loss of cognitive function and, ultimately, a diagnosis of dementia. For instance, the decline of cognitive functions, loss of short-term memory, and decreased bodily functions in the late stage of the disease are common characteristics of AD. It is believed that such interference is associated with neuronal death, brain atrophy, and overall cognitive decline because of toxic protein aggregates throughout the brain.

Diagnosing neurodegenerative diseases nowadays relies on biological markers, called *biomarkers*, to complement clinical cognitive assessments [78, 195]. Biomarkers usually indicate specific in vivo alterations [87, 88], either physical or molecular abnormality. Clinicians usually rely on brain atrophy to examine the loss of brain matter due to cell death and to correlate the disease stage with mental and physical difficulties [92]. They detect atrophy through autopsy or structural brain magnetic resonance imaging (MRI). For instance, Figure 1.1 shows the structural MRI of brain-shape alterations over two years of different patients. Despite the advantage of observing in vivo alterations, this biomarker cannot characterise a specific histopathology of the disease. Together with physical biomarkers, clinicians also use molecular biomarkers such as toxic protein aggregates (also known as misfolded isomers) which are the disease agents accumulating in the brain. The well-known proteins in neurodegenerative

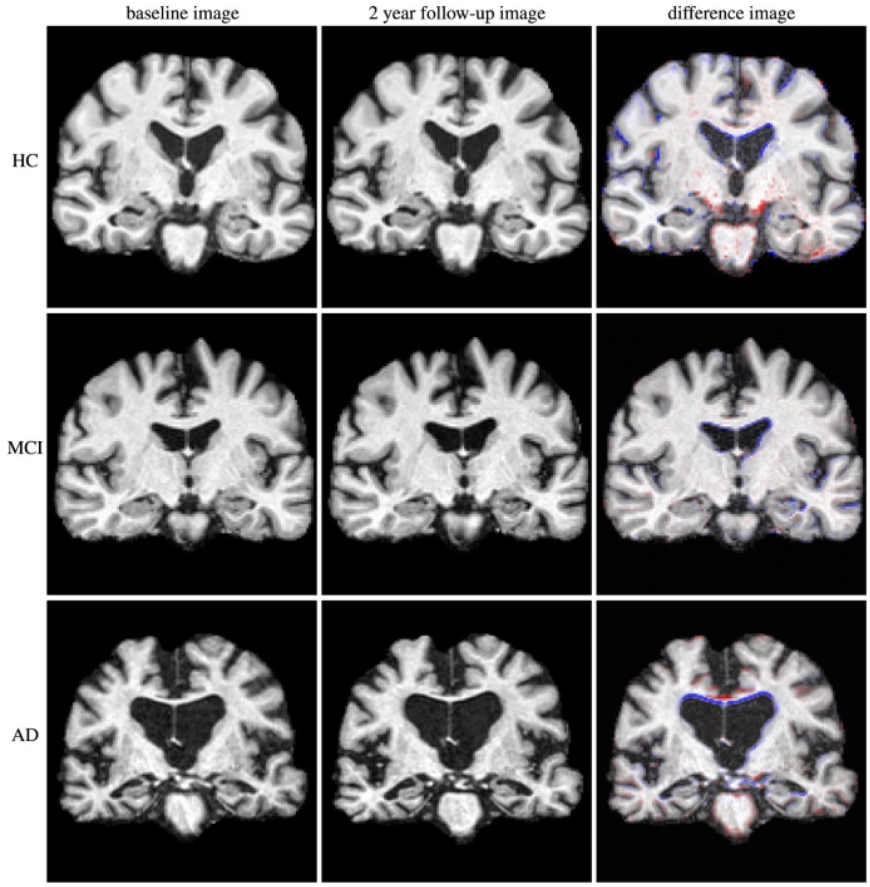


Figure 1.1: Brain MRI imaging. Comparison of brain shape alterations over two years of a healthy patient (HC; 84.8 years at baseline), a patient with mild cognitive impairment (MCI; 71.8 years at baseline), and a patient with AD (AD; 77.5 years at baseline). After two years of each brain patient, the subtle alterations are indicated on the right column; adopted from Ledig et al. [110].

diseases are amyloid-β (Aβ) and tau protein (τP) in AD and α-synuclein (αS) in PD. Clinicians examine biomarkers through either cerebrospinal fluid (CSF) or positron emission tomography (PET) imaging. CSF provides pathological information in the early progression of a neurodegenerative disease [15], for example, Aβ40 and Aβ42[1] concentration which is an indicator for AD. Meanwhile, PET is an imaging technique that provides information about physiological and biochemical activities of the brain [92] (see Figure 1.2 for example). Moreover, PET can also be used to record brain function. For instance, 18F-fluorodeoxyglucose (FDG) PET imaging measures brain activity dysregulation. The use of biomarkers has been reported to be better at

[1]Aβ40 and Aβ42 are major isoforms of Aβ. Aβ42 contains two additional C-terminal residues compared to Aβ40. This isoform is the predominant component of amyloid plaques, while Aβ40 is dominant in CSF [66].

predicting the disease prevalence at different ages [86] and, especially, to distinguish the MCI and AD stages [78].

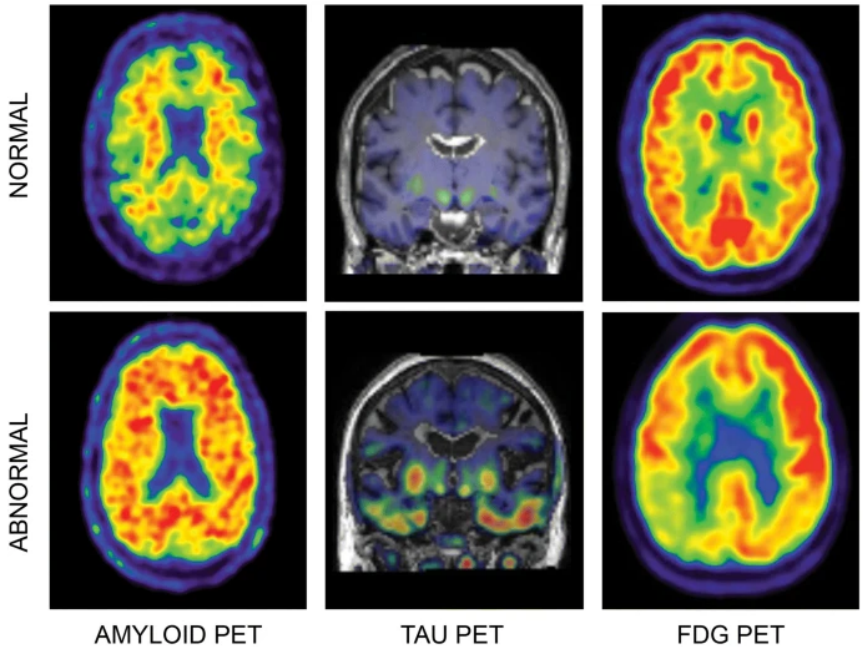


Figure 1.2: PET imaging. Normal brain (top) and abnormal brain (bottom) PET imaging in three different scans. Aβ tracer scan (left), τP tracer scan (middle), and 18F-fluorodeoxyglucose (FDG) tracer scan (right). Warmer colors (yellow, orange, and red) show more tracer binding indicating more Aβ deposits, τP tangles, and better brain activities in FDG scan, respectively; adopted from Ten et al. [169].

The use of biomarkers is critical in order to explain the disease cascade and to enable the diagnosis of neurodegenerative disease as early as possible. The amyloid PET scan for AD, for instance, has become one of the diagnostic requirements because it will show an increased cortical concentration of the tracer in regions of amyloid-aggregate deposition [178]. If not accessible, the CSF is an alternative tool to assess the disease progression to measure the concentration abnormality. In order to understand the disease cascade, the evolution of different biomarkers has been theoretically modelled into abnormality curves that increase over time towards the severe stage of the disease [87]. Figure 1.3 illustrates the sequential evolution of AD biomarkers where the concept emphasises the ordering of Aβ and τP detected by CSF and PET imaging at a given detection threshold. As the other biomarkers (i.e. FDG PET and MRI analysis) rise, cognitive impairment becomes evident, with a range of possible cognitive responses that vary amongst individuals. The authors suggest that this biomarker model reflects in vivo functional changes of AD but not autopsy-based

pathological histology [87].

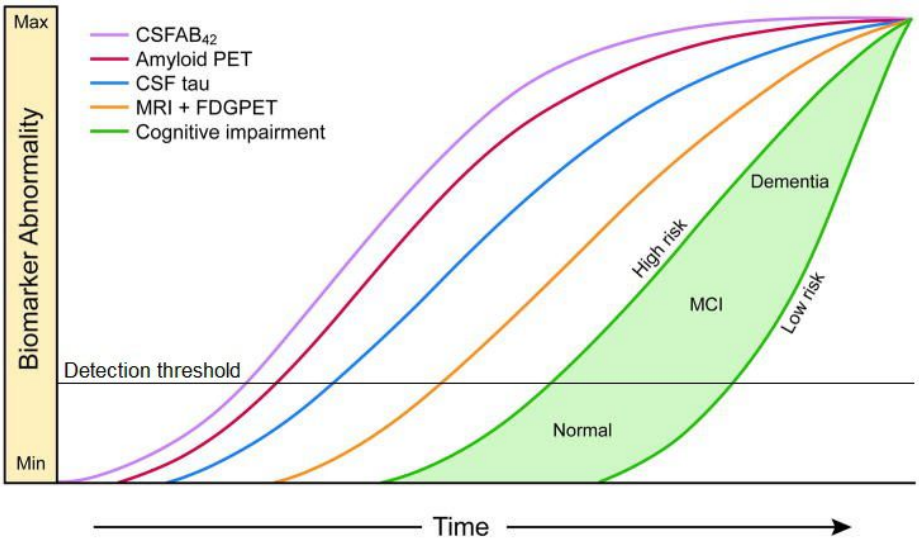


Figure 1.3: Conceptual model of biomarker evolution for AD known as Jack's curves. Aβ deposition has been found to occur in the early stages followed by τP deposition followed by MRI atrophy and decrease in cognitive functions; adapted from Jack et al. [87].

Despite the advantages and possibilities of biomarkers, the practical and cost constraints of clinical AD experiments, most especially in humans, have led to an interest in developing mathematical models of pathology evolution. Mathematical modelling provides the essential foundation to describe large-scale features of neurodegenerative disease through the quantification of pathology. It allows for integrating different hypotheses and coupling systematically various contributing factors in order to reveal the complex mechanism of biomarker progression in the human brain.

1.2 Scope of the thesis

In this thesis, I study the in vivo dynamic of toxic τP as an AD biomarker. Toxic τPs have been observed to traverse throughout the brain systematically and hierarchically. According to the seminal work of Braak and Braak [20], AD severity can be categorised into six different stages in which each stage is associated with toxic τP load at a specific brain region of interest. Recent works have suggested that the transentorhinal region (or entorhinal in general) acts as a seeding location [180] and transport of toxic τP occurs through communicating neurons [44, 180]. The key question of this thesis is to understand the spatiotemporal evolution of toxic tau proteins in the human brain through axonal links.

The concept of prion-like growth and toxic protein spreading has become a central proposal of the underlying neurodegeneration mechanisms. The disease etiology indicates a corruption chain of large toxic molecules in the brain [6, 49, 50, 184, 187]. The amplification process is similar to the one found in prion disease pathogenic conformation [141] where the spreading of toxic proteins occurs through intracellular transport mechanisms [21] and extracellular space by secretion or damage of the host cell [161]. To address the proposed question, I develop mathematical network models based on the prion hypothesis coupled with a linear transport principle. In this thesis, I use computational and asymptotic approaches to capture the hallmark features of this process and extract information about toxic τP propagation

1.3 Thesis organisation

The work in this thesis is organised as follows. I review the literature regarding toxic proteins in neurodegenerative disease and mathematical modelling of neurodegenerative disease in Chapter 2. In this chapter, I also determine our global and specific objectives. Chapter 3 discusses the staging problem. I propose computational methods called *braid diagram* and *braid surface* to reveal staging within the network-organised reaction-diffusion. Chapter 4 consists of the arrival-time problem. We propose methods to estimate arrival times and apply them to neurodegenerative diseases and τP propagation on undirected networks in Alzheimer's disease. Chapter 5 consists of the study of heterogeneous systems, while Chapter 6 discusses the dynamics of directed networks. I propose a study with respect to staging and arrival time for both chapters to reveal the effect of directional bias and regional capacities on the resulting dynamics. Finally, I conclude all work and propose potential future directions in Chapter 7.

Chapter 2

State of The Art

2.1 Evidence regarding Alzheimer's disease

A hallmark of neurodegenerative diseases is the large presence of toxic protein accumulation throughout the brain. The proteins that present under healthy conditions misfold, replicate, and spread in a pathological state and, eventually, bring harm to the brain neurons [60, 186]. The most common neurodegenerative disease is AD and it involves Aβ and τP over the disease progression. In this section, we discuss evidence of AD key biomarkers that characterises the disease evolution.

Both proteins have beneficial roles in the human brain but become harmful due to certain factors. Aβ is produced from the proteolytic cleavage of the amyloid precursor protein (APP), a large membrane protein that is ubiquitously expressed in healthy brains and predominantly plays a vital role in neural growth and repair [73, 133]. If an APP cleavage is first processed by an enzyme called α-secretase, the resulting protein is not harmful and prevents Aβ formation. An APP is processed instead by an enzyme called β- and then γ-secretase results in toxic and misfolded Aβ. The toxic peptides can accumulate over time to produce senile plaques with neurotoxic effects. τP is mostly found in the axons. It is encoded by the microtubule-associated-protein-tau (MAPT) gene, located on chromosome 17 [4]. The protein's crucial role is to stabilise the microtubules that provide a skeleton structure for the delivery of substances from the cell body down the synapses by interacting with tubulin [67, 167]. The other way to stabilise microtubules is through the phosphorylation of τP. However, pathological conditions can increase its phosphorylation process and reduce its binding ability to microtubules. It results in skeleton destabilisation and toxic fragment release from the conformation [67]. Similar to Aβ, toxic τP fragments bind together and create toxic and misfolded neurofibrillary tangles (NFTs). Once their toxic and misfolding isoforms accumulate in the brain, it ultimately leads to neuronal death.

2.1.1 Post-mortem evidence of protein biomarker

Aβ and τP, initially observed through autopsy, are the standard molecular indicators to study disease progression. Cross-sectional analyses of post-mortem human brains revealed a characteristic progression of Aβ plaques and a highly predicted appearance of neurofibrillary tangles. The clinical course of the disease based on the biomarkers can be subdivided into a number of stages for better recognition and control of the disease progress [18, 170].

A seminal autopsy study has revealed that the evolution of Aβ deposition in the brain allows the distinction of five phases [170]. The overall progression indicates that senile plaques develop first in the neocortex, continue to the allocortex, and eventually to the subcortex (Figure 2.1). This sequence in which the regions of the brain are involved suggests that the deposition occurs through a successive process. The neocortex is always the first region to develop Aβ deposits. It is observed that plaques disperse focally into small groups in the grey matter. In phase 2, Aβ deposits appear in the allocortex, with possibly, additional plaques in the amygdala, entorhinal cortex, and cingulate gyrus. Phase 3 is characterised by the development of plaques in the subcortical regions such as the caudate, putamen, thalamus, hypothalamus, and white matter. Additional Aβ deposits in phase 4 appear in the brainstem nuclei and substantia nigra, where it is related to the control, cognitive executive functions, and emotional limbic activity [160]. Phase 5 is predominantly characterised by the plaque formation in the cerebellum, frequently in the molecular layer and rarely in the cerebellar granular layer. It is believed that Aβ deposition in the brain spreads in an anterograde direction. This remark indicates that Aβ traverses from regions already exhibiting plaques into regions that receive neuronal input from these regions. For example, Aβ in the neocortex becomes involved in the amygdala in phase 2 and the pontine nuclei in phase 5 [170]. The finding of the amyloid phases suggests that AD is a disease in which large areas of the brain develop toxic plaques (phases 1 to 3) before clinical symptoms appear. In addition, these phases are correlated significantly with the evolution of neurofibrillary tangles. The deposit progression corresponds to functionally and anatomically connected brain regions [27, 170, 183].

Earlier than the finding of the structured Aβ evolution in the brain, Braak and Braak proposed a crucial AD disease staging according to the hierarchical τP spreading [18]. They suggested that hyperphosphorylated τP load can be distinguished into six stages within three key areas, i.e. the transentorhinal, the limbic, and the isocortical areas (Figure 2.2). The distribution pattern exhibits significant differentiation of neuropathology. Unlike Aβ plaques, τP tangles deposited in the neocortex

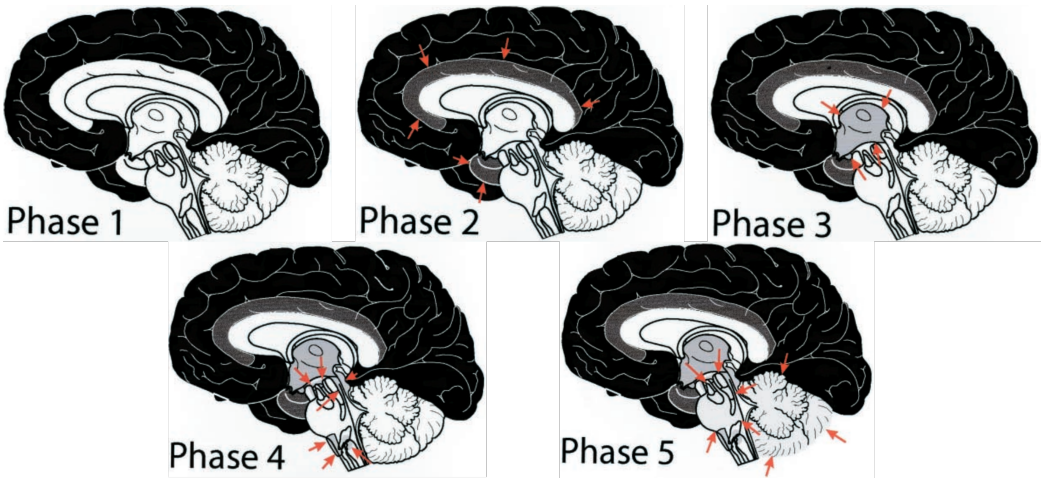


Figure 2.1: Aβ propagation. Phase 1 is shown by neocortical Aβ deposits (Neocortex: black). Phase 2 shows additional allocortical Aβ deposits (red arrows), phase 3 additional Aβ deposits in diencephalic nuclei (red arrows), phase 4 additional Aβ deposits in distinct brainstem nuclei (red arrows), and phase 5 in the cerebellum and additional brainstem nuclei (red arrows); adapted from Thal et al. [170].

do not accumulate erratically [18, 41, 171]. Stage I is characterised by the appearance of neurofibrillary tangles in the locus coeruleus and entorhinal areas. Stage II is the aggravation of the first stage and τP tangles start to affect the hippocampal area. Stages I and II are often referred to the *transentorhinal stages* and correspond to pre-clinical AD, which initiates τP deposition and shows no symptoms. In stage III, neurofibrillary tangles appear in the parahippocampal region and the limbic area such as the amygdala and thalamus. It is frequently observed in this stage that tangles affect the dendritic area in the brain regions and convert into an extraneuronal structure called the *ghost tangle*. Stage IV is characterised by a large number of ghost tangles in the regions associated with stages I and II. A further feature of this stage is that neurofibrillary tangles mildly affect the isocortex and large neurons in the basal portions of the putamen and the accumbens nucleus. Stages III and IV are known as the *limbic stages* and correspond to the clinical AD that is showing a cognitive decline in AD patients. The last two stages, often called the *isocortical stages*, show severe changes of neurofibrillary tangles in large area of the brain. The main feature of stage V is that the isocortex is severely affected, while stage VI shows even more changes and considerable loss of neuronal cells. A special feature of the last stage is the presence of neurofibrillary tangles in the primary sensory area that contains soma-devoid nerves [44]. The isocortical stages correspond to severe AD. A recent study suggests that τP tangles and soluble seeds exist prior to pathology and other

regions that are synaptically connected to the seeding location [44]. This consistent presence, density, and distribution pattern of τP tangles is therefore known as the *Braak stages*.

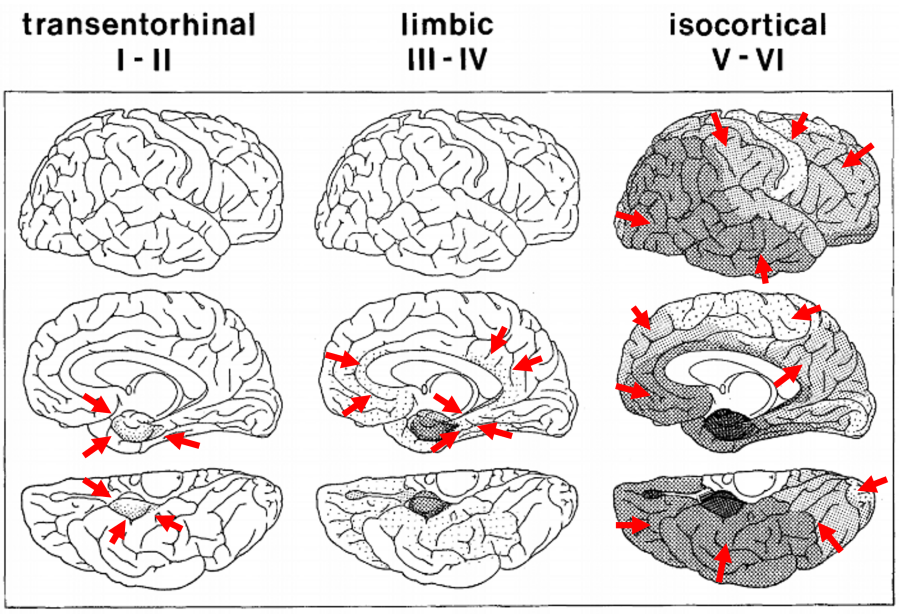


Figure 2.2: Distribution pattern of hyperphosphorylated τP in AD. Alterations in the transentorhinal region (red arrows) occur in stage I-II (left), severe involvements in the entorhinal region and transentorhinal region indicate impairment in the limbic area (red arrows) in stage III-IV (middle), and increasing intensity in the isocortical (red arrows) region occurs in stage V-VI. Darker shades indicate increasing severity in τP load; adapted from Braak and Braak [18].

2.1.2 Further evidence of protein biomarker

Biomarkers these days are often identified through CSF and/or PET scans instead of post-mortem. The two methods, CSF analysis and PET scans, can provide pathological information in the early progression of neurodegenerative disease [15]. Investigating biomarker abnormality accurately is important, especially during the onset of AD. PET imaging, particularly, allows studying in vivo progression of neurodegenerative diseases.

Amyloid deposition can be detected through CSF. It is possible because amyloid plaques directly contact the extracellular space of the brain. Any biochemical changes or abnormal amyloid concentration in the brain are therefore reflected in the CSF [15]. CSF Aβ42 is a basic measure of fibrillar Aβ42 and plaques in the diseased brain. It has been shown that levels of Aβ42 in post-mortem CSF have a relation with plaque

load at autopsy [164, 168]. Aβ42 concentration in the CSF has a negative correlation with the amyloid deposition in the brain shown by an increase of plaques in the brain reduces the Aβ42 concentration detected in the CSF [168]. However, there is a bias found in the CSF Aβ42 that it is tightly correlated with amyloid retention in brain regions adjacent to CSF spaces [65, 92].

Amyloid PET imaging is initially purposed as an in vivo surrogate for Aβ pathology. It is used to detect cerebral senile plaques by giving signals to confirm amyloid prions in the brain [92, 179]. Different studies have consistently reported that CSF Aβ42 has a good correlation with Pittsburgh Compound B (PIB)-PET binding tracer and 11CPIB binding tracer [15]. Amyloid PET is believed to be a more suitable measure to detect changes in the brain plaque load compared to the CSF Aβ42 because it strongly and directly correlates with the aggregates [82]. Using the PET scans, we are allowed to identify and quantify amyloid load regionally. However, it has been suggested that amyloid PET is not a good marker during the clinical stage of AD [92]. The scans have a beneficial use in separating cognitively normal and impaired elders. It is believed that the detection of amyloid load during the clinically invisible stage can provide insights into the beginning of the disease onset.

An example of extensive use of amyloid PET is a study [31] to determine the sequence of amyloid propagation in the brain in order to check whether Aβ in PET scans follows a stereotypical pattern of propagation as in [18, 170]. The authors use Z-scores derived from the PET scores and involvement percentage of amyloid involvement frequency from patients after separating between positive and negative amyloid patients. The data suggest that the range of percentage involvement of all neocortical regions is small and there is no difference in involvement frequency within the neocortical regions. They argue that small amyloid deposits in multiple neocortical areas may spread simultaneously through multiple schemes, which is aligned with [18, 170]. This observation, however, indicates that amyloid PET is unsuitable for determining AD severity, especially if it depends on the cortical spreading pattern of amyloid pathology which is unpredictable.

τP that is the primary component of neurofibrillary tangles has been assessed through the CSF as the levels of this protein show an important pathologic feature of AD [15]. CSF τP has been frequently reported to have a positive correlation with the size of the damaged tissue and a negative correlation with clinical outcome [74, 129]. A quick development from mild cognitive impairment (MCI) to AD is reflected by a high CSF τP [16]. Also, high CSF τP can be used to identify quick cognitive decline in AD patients [149]. It has been reported that CSF τP increases significantly as

Aβ deposition increases. It is believed that the CSF τP level depends on fragments released from tau-bearing neurons [168]. These findings support the hypothesis that the CSF τP level reflects τP phosphorylation and the formation of tangles in the central nervous system. However, CSF τP is not useful in distinguishing pre-clinical AD and AD patients [15].

τP deposits which aggregate intracellularly are basically difficult to access in vivo until PET scans are used. Similar to amyloid PET, tau PET provides signals that associate with tau prions. It has been reported that [18F]AV1451 tau PET can be used to differentiate between pre-clinical AD and AD patients since signals from AD are higher [93, 111]. It is suggested that tau PET is useful as a stage marker that reflects disease progression since there is a significant signal elevation from the cognitively normal to the MCI state [117]. It is also reported that tau PET is more closely related to neurodegeneration biomarkers such as atrophy and metabolism disruption than amyloid presence. Despite its potential for studying AD, studies report that the low spatial resolution of PET and the off-target binding of [18F]AV-1451 limit tau PET accuracy [148].

Assessment of tau PET has shown a regional τP pathological progression in the brain [31]. The authors use a similar process to the amyloid identification in the same paper where Z-scores derived from the PET scores and involvement percentage of tau involvement frequency from patients after separating between positive and negative tau patients. An increasing trend for τP binding has been observed when assessed between different stages. However, image-based τP staging does not show the exact Braak staging. For example, τP binding in the hippocampal region is not found at Braak stage II. This finding may be due to variable off-target binding in the choroid plexus. Furthermore, lingual and insular regions that are originally classified into stages III and IV are characterised into stage V. Although tau PET allows the identification of image-based staging, this finding indicates a notable difference between post-mortem and image-based studies. However, tau PET could still be a useful biomarker for the clinical and pathological progression of AD.

2.1.3 Evidence of brain atrophy

Progressive cerebral atrophy is another hallmark of neurodegeneration associated with toxic protein aggregates. Brain atrophy is defined as a loss of brain matter due to cell death and decreased brain volume [92]. The major contributors are dendritic and neuronal loss. The major changes occur in the temporal lobe while the frontal and temporal cortices have enlarged sulcal spaces and lateral ventricles [43]. MRI

studies reveal that the sequence of atrophy evolution fits the histopathological of neurofibrillary tangles [92] while amyloid deposition does not correlate with neuronal loss [39, 61, 134]. Both neuronal loss and neurofibrillary tangles increase as the disease progresses [61, 134]. From post-mortem studies as shown in Figure 1.1, AD brains have a thinner cortex compared to MCI and cognitively normal subjects, principally in bilateral, frontal, parietal, temporal, and occipital lobes [45]. Atrophy in MCI subjects is identified to resemble healthy elderly and shows a significant difference in the transition from MCI to AD [156]. It is suggested that neuronal loss in the superior temporal sulcus can lead to cognitive decline in AD [61] while cortical thickness negatively correlates with dementia severity in AD [45].

Changes in regional volume reflect the atrophy of the associated region. Studies of regional MRI volumes suggest that volume alterations are closely correlated with neuronal counts from post-mortem analysis [92]. An extensive study on brain volume through MRI scans has captured a predominant feature of atrophy found in autopsy [31]. Volume reduction of cortical grey matter is most frequently observed in the medial temporal regions such as the hippocampus, amygdala, entorhinal, and parahippocampal cortices with the largest alteration occurring in hippocampal volume for around 23% difference between cognitively normal and AD subjects. When assessment is done according to the Braak staging, it has been found that brains in stages I-IV do not show a cortical thinning process. The apparent neurodegeneration happens to a few regions in stage V such as the entorhinal, parahippocampal, and inferior temporal cortices. Cortical thinning becomes clearly apparent in many regions in stage VI. Although the regional volume spreading pattern is not that significant to protein biomarkers, it shows a similar propagation to τP accumulation. This finding is aligned with previous studies [43, 45, 61, 92, 134].

2.1.4 Insight from evidence

Studies conducted on Alzheimer's disease provide a lot of evidence regarding the disease's progression. It is indeed toxic protein aggregation is the hallmark of such a disorder. However, we highlight three notable features observed regarding the pattern of accumulation in the human brain.

First, toxic protein aggregates accumulate gradually from region to region. Clinicians observe that aggregates initially start from a critical amount of seeds in one or more sources in the brain and propagate systematically and hierarchically. The finding of regional spreading allows clinicians to classify disease severity into distinct stages and each stage is associated with one or more brain regions [18, 170]. Both

toxic Aβ and τP exhibit this notable feature. However, tau tangles show clearer isocortical staging than Aβ plaques. We can study the spatiotemporal evolution of toxic proteins from this feature. It reduces complexity by dividing the brain into smaller patches and, at the same time, enabling concentration differentiation over time for each small patch.

Second, PET imaging demonstrates in vivo toxic protein propagation. The development of advanced imaging techniques and tracers allows for studying the regional spreading through signals as surrogate concentrations. Additionally, it allows for studying pathological evolution during disease progression. Clinicians have shown that PET signals consistently correlate with protein aggregates and clinical status [82, 93, 111]. We can use PET images as a tool to study toxic protein spatiotemporal evolution instead of analysing CSF. We can extract regional information from PET signals but not from CSF since it only provides temporal disease progression. Despite its benefits, we also note that PET imaging suffers from off-binding [148] which reduces accuracy and shows different patterns of regional spreading [31]. Therefore, we may need to handle PET signals carefully when comparing them against the simulation outcomes.

Third, tau protein, compared to Aβ, shows a structured progression in disease development. Staging of τP is more apparent and clear to classify. A study of τP reveals that the protein strongly correlates with brain damage [117]. PET signals of τP can separate the healthy and diseased brains which allow differentiating significant signal elevation between brains.

We conclude that studying the evolution of tau proteins is essential to understanding the spatiotemporal evolution of neurodegenerative disease in the brain.

2.2 Proposed mechanisms in Alzheimer's disease

Fundamental evidence from AD studies has led to important hypotheses for understanding of the disease's etiology. The process and pattern that emerge in AD implicate intricate co-mechanism between toxic proteins, neuronal transport, and synaptic activities within the brain. Therefore, we need to break down the individual knowledge that constitutes the etiology.

2.2.1 Amyloid cascade hypothesis

Alzheimer's disease is characterised by various biomarkers with three main well-known abnormalities in the brain, i.e. amyloid senile plaques, neurofibrillary tangles, and

vascular damage, which result in neuronal loss [70, 71]. According to the amyloid cascade hypothesis, the pathogenic mutation in the Aβ precursor protein (APP) is the primary event of AD progression. The hypothesis suggests that Aβ deposition is the initial occurrence of AD in humans, especially those who suffer from Down Syndrome [115, 147]. Consistent with this hypothesis, the process likely progresses to initiate excess τP phosphorylation and tangle formation, ultimately leading to neuronal death. This so-called *amyloid cascade hypothesis*, proposed by Hardy and Higgins [70, 71], has been the basis to most major AD studies and drug development.

Studies conducted around the hypothesis have been mainly focused on understanding amyloid formation and senile plaques to reveal their role in developing such diseases. It has been found that amyloid deposition may occur due to an amyloid precursor protein (APP) gene mutation. Unfortunately, the major variants lack a substance that can inhibit excessive amyloid production in the human brain [99, 108]. Observations suggest that amyloid plaques that come from the fragments of APP diffuse throughout the brain [98] and are found as the only form of senile plaques as commonly discovered in Down syndrome patients [115, 147]. These diffuse plaques are distributed widely in the central nervous system and can evolve into neuritic plaques in specific brain locations [28, 127]. Plaques that contain glutamine mutation for glutamic acid attach to the walls of cerebral blood vessels and can lead to neuronal loss, ultimately, cerebral haemorrhage.

Amyloid plaques commonly comprise β-polypeptide fibrils as waste neurites surround the core with astrocytes and microglia. It has been observed that axon-and-dendrite debris surrounding the amyloid core contain paired helical filaments (PHFs), also a part of neurofibrillary tangle structure [105]. Furthermore, PHFs have been observed to contain Aβ fragments and tau molecules. Nevertheless, it is not clear what the main components of PHFs are and the underlying binding mechanism due to possible secondary amyloid deposition in the surface of tau tangles or absorbed into the tangles [193, 194]. Despite the controversies related to PHF elements, it is believed that plaques and tangles do not emerge independently. The discovery of APP suggests that amyloid plaques act as the primary lesion that precedes tau tangle formation. However, they disrupt neuronal cells that contribute to neuronal function deficiencies and, eventually, the syndrome of dementia [70, 71].

2.2.2 Tau hypothesis

The previous hypothesis predominantly emphasises the perspective of Aβ plaque as the trigger of the onset of AD. However, further investigations on non-demented

patients and mouse models indicate that amyloid deposition has no direct implication with the beginning of the disease and is commonly observed amongst healthy ageing elders. This counter-intuitive statement was supported by evidence that AD patients may have very few amyloid deposits and healthy patients may develop abundant amyloid deposits comparable to demented patients [46, 113]. The existence of Aβ oligomers and fibrils in amyloid plaques in mouse brains did not lead to neuronal loss or cognitive impairment [106, 107]. Furthermore, mouse models with Aβ deposits did not show NFT formation and neuron death [24]. These observations challenge the hypothesis that amyloid deposition is AD's primary lesion.

One of the AD characteristics is the abnormality of tau tangle deposition in the brain. Studies have shown that tau lesions start earlier than Aβ deposition [19, 93]. The idea, called the *tau hypothesis* [59, 97, 121, 123], suggests that the progression of AD is strongly associated with tau pathology which causes extensive damage in neurons through impaired cytoskeletal, signalling, and damage to the transport system in the brain. The spreading of tau pathology has been observed to strongly correlate with cognitive impairment and clinical symptoms which studies using PET scans amplify the statement by showing a resemblance between tracer and neurodegeneration patterns [13, 25, 26, 128, 151, 153]. The tau hypothesis emphasises that the principle causative substance of AD is τP. It has received much attention among neuroscientists in the last five years.

Tau pathology in AD is observed to spread hierarchically in brains identified into six Braak and Braak stages [18] as discussed in section 2.1.1. Tau aggregation is also seen in other neurodegenerative disorders such as frontotemporal dementia, Parkinson's, and progressive supranuclear palsy. A common feature found in these diseases is tau gene mutation [3] inducing hyperphosphorylation to impair the ability of τP to bind microtubules [72, 79, 139, 163, 162, 196]. Tau fragments then bind and form fibrils, tangles, and ultimately neurofibrillary tangles [11, 192]. These observations suggest that tau accumulation is a cause of neurodegeneration rather than amyloid plaques, although the initial flame of AD is still unclear.

2.2.3 The prion-like hypothesis

As neurodegeneration progresses, an increasing abnormality of biomarkers indicates load accumulation of protein causing neurodegeneration. Studies on Alzheimer's disease have suggested that the disease etiology indicates a corruption chain of healthy molecules into misfolded forms which involve molecule aggregation and fragmentation of large toxic molecules into smaller ones [6, 49, 50, 184, 187]. This amplification

process is similar to the one found in prion disease pathogenic conformation [141], where the term *prion* refers to a type of protein that can cause healthy proteins to fold abnormally. This mechanism is called the *prion hypothesis*.

Prusiner initially proposed the mechanism of prion infectivity in 1991 as an exponential replication in which healthy protein conversion into abnormal isoform occurred [141]. Protein labelled as prion has two forms: healthy isoform, presented within the cell, and toxic isoform, which contains a specific viral nucleic acid. Prusiner proposed that these two isoforms combine to produce a heterodimer substance and fragment into two toxic molecules. In the next cycle, these molecules combine with two healthy molecules yielding four toxic molecules. The replication mechanism occurs repetitively, increasing toxic prions exponentially through this heterodimeric intermediate process and possibly becoming autocatalytic.

Experiments at the molecular level have demonstrated that in AD, Aβ protein shares similar features and proliferation kinetic to prion [6, 184]. Aβ extract injection on transgenic mice demonstrated that neuritic changes and neuroinflammation can form after the introduction of Aβ seeding on the brain [94]. Furthermore, this protein can traverse from neuron to neuron in vitro and progress along the brain's anatomical pathways in vivo. In addition, Aβ can form different strains and sizes, but the small and soluble ones are the most likely to propagate.

Tau aggregation in the brain appears to comply also with the prion-like hypothesis. It is often considered as the outcome of Aβ abnormality cascade [70, 71], but further studies indicates τP may be the torch of AD [18, 19]. Similar to Aβ experiments, injecting τP extracts into healthy transgenic mice induces τP seed release and accumulation. A distinct identified feature of Aβ is that the proliferation of τP seed can occur not only in transgenic mice but also in wild-type mice [184]. Also, τP has a range of sizes, but the most toxic ones are believed to be the small and soluble oligomers.

2.2.4 Spreading of toxic proteins in the brain

Toxic proteins, especially in their soluble form, can move between connected regions in the brain and imprint some general striking features. A notable trait of toxic protein causing neurodegenerative disease is that the seeds start from an initial seeding location [68]. Disease progression is inevitable after initial seeding. However, there is a clinically silent phase during which toxic proteins grow and spread, similar to the progression of prion disease [116, 142]. Another trait from post-mortem

analysis and animal model studies on Alzheimer's disease, and other neurodegenerative diseases in general, is a particular progression of repeatable and predictable patterns [18, 34, 95, 170]. In case of Aβ and τP in AD, Amyloid-β deposits appear initially in the neocortex and spread into the allocortex and the subcortex [170] and τP occurs first in the locus coeruleus and transentorhinal layer and then spread into the amygdala and interconnected neocortical brain regions [18]. Toxic protein seeds can undergo regional transmission [44, 182] through the self-propagation mechanism [95].

Toxic proteins can spread, in general, through either axonal transport or extracellular diffusion [187]. Axonal transport is believed to correspond to neuronal activity [154]. In [125], in vitro misfolded seed proteins demonstrate a direct neuronal exchange through the axonal pathway. It is observed that seeds can move bidirectionally along axons and spread between synaptically connected neighbours [44, 109, 190]. Diffusion is usually understood to be associated with neuronal proximity [54]. Different studies in [124, 135, 138] have demonstrated that seeds can move into the extracellular space through several possible ways. They can diffuse through secretion and cell-to-cell endocytosis [161]. It is therefore suggested that axonal transport is the major mechanism in the spread of seeds [95].

Although Aβ deposition is formed extracellularly and in direct contact with the extracellular space, its seed form traverses between synaptically connected regions. Evidence shows that Aβ expands anterogradely into regions that receive neuronal projections from regions already exhibiting Aβ [170]. In [125], Aβ oligomers are observed to transmit neuron to neuron through direct neuritic connections and intracellular oligomers play a role in neurodegeneration instead of extracellular form. This event is shown by internal neurite damage and gradual endosomal leakage because of accumulated Aβ after introducing seeds. It is still debated whether the intraneuronal Aβ aggregates are produced from Aβ secreted extracellularly, later taken up by cells or formed directly from Aβ produced intracellularly [69]. In [53], the formation of Aβ amyloid plaques has been observed in which initially soluble and extracellular Aβ peptide becomes internalised. Once aggregated, internalised Aβ oligomers can increase the accumulation of newly synthesised Aβ and act as seeds for intracellular aggregation [125]. Ultimately, cells die and all intracellular components including Aβ oligomers are released into the extracellular space. In addition, Aβ is reported to be transported in exosome through endocytosis and exocytosis processes from in vitro cell-to-cell transfer [125]. Also, there is evidence that Aβ seeds can travel from the periphery to the brain as observed in a mouse model [94] which is postulated to occur

through the vascular system [39]. Findings supporting this proposed mechanism are consistent with substantial evidence from AD brain tissues and AD animal models.

τP is believed to move along the axonal pathways when fragments detach from the conformational structure. τP can detach because of hyperphosphorylation that reduces its binding affinity to the tubules [67]. When these molecules bind with each other, the aggregates act as toxic seeds and move either towards the soma or axonal distal end [63]. Smaller and soluble aggregates move to the synapses [44, 109]. Empirical studies indicate that the anterograde movement of soluble seeds towards the synapses is a microtubule-dependent transport through piggy-backing free microtubule fragments or fast stop-and-go motion [21, 109]. Meanwhile, larger oligomers move retrogradely towards neuronal soma and accumulate there. The accumulation cluster of τP in soma and neurites has been identified in cell cultures by decreasing concentration gradient from the body cell to the axon end [109]. In addition, it has been observed that motion of τP along the axon inhibits the anterograde transport of organelles and vesicles. In [44], an experiment on a larger scale shows that τP seeds are detected not only in the infected regions but also in the subsequent non-infected region along the Braak pathway. τP seeds and aggregates have been observed post-mortem within the synapses and along the axon of soma-devoid nerves. These findings show that τP traverses via axons and synaptically connected regions.

Different studies reported that τP molecules also traverse extracellularly. The protein is believed to have particular functions in the extracellular space such as to increase the electrical activity of primary cortical neurons [22, 191]. Different studies have mentioned that intracellular τP aggregates exist in the extracellular space through secretion because of clearance mechanism or neuron degeneration in both physiological and pathological conditions [137, 190]. Similar to Aβ, τP is released as either free τP molecules or exosome mechanism inside vesicles [137, 138]. Upon axonal or somatodendritic degeneration, some released fragments could be internalised by other neurons through endocytosis, then be transported anterogradely and retrogradely throughout different regions, accumulate, and, ultimately, reduce the cellular clearance process in the brain [161, 190]. In addition, extracellular aggregates internalised at the axonal terminal can spread retrogradely toward soma [190] and interrupt the anterograde transport of organelles and vesicles [109]. Studies that support this mechanism show that the extracellular τP fibrils are engulfed into the cells and start aggregating soluble τP in culture cell [54] and in a mouse model [34]. It is also observed that trans-neuronal propagation takes place when τP fibrils are introduced into the brain in vivo [34, 81].

2.2.5 Highlight on disease mechanism

Hypotheses and theories in AD mainly come from the studies on the amyloid cascade hypothesis. Accordingly, it is believed that Aβ is the initial trigger of the disease. However, further studies unveil different ideas regarding the cause of the disease and the transport mechanism. Here, we summarise two key factors of toxic protein expansion and transport in the brain.

First, toxic proteins related to neurodegenerative disorders follow prion characteristics. The process of nucleation, growth, and spreading is similar to prions where proteins proliferate silently and slowly in the beginning and eventually become fatal. Studies of different neurodegenerative diseases produce consistent evidence regarding this prion-like self-replication [35, 60, 95, 96]. This mechanism is called the *prion-like hypothesis* and it is central for our model of toxic protein evolution.

Second, the key spreading of toxic protein in the brain is intracellular transport. Evidence suggests that τP molecules particularly reside in the axons and move between neurons through the axonal channels [44, 109]. There are several proposed mechanisms for explaining τP intracellular movement, yet empirical studies suggest diffusion as the main process [54, 109]. Another experiment also shows that Aβ aggregate initially accumulates intracellularly and moves neuron to neuron [125]. Despite the controversy, the intracellular process actually becomes crucial in extracellular Aβ plaque formation. Accordingly, we take into account intracellular transport as another fundamental part of our model of toxic protein transport.

2.3 Mathematical modelling review

Post-mortem studies or animal model studies on Alzheimer's disease and other neurodegenerative diseases, in general, have demonstrated different mechanisms of neurodegeneration at the molecular level. However, these processes within the brain remain unclear and assessing pathology in vivo is challenging. To fully understand how neurodegeneration takes place inside the human brain, mathematical and computational modelling become essential and beneficial.

We now need to take into account the prion-like paradigm and intracellular transport in our mathematical model. First, the models of prion diseases in the past shape the understanding of how prion-like toxic proteins aggregate in the brain. We learn from the prion paradigm that a consistent model must elucidate: (i) the inevitable disease progression after immediate inoculation and (ii) the slow disease progression at the early stage [47, 90, 116, 141, 140]. These models agree that the autocatalytic

process is the common feature of the prionic behaviour. Second, a major contributing factor to neurodegeneration is the intracellular transport of toxic protein aggregates along the axons. We note that various mathematical models mostly use the diffusive paradigm to explain prion-like protein movement or delivery in axons [21, 109, 174].

The autocatalytic process and diffusion are key in modelling toxic protein transport in neurodegenerative diseases. Some notable studies have provided a vital understanding of the underlying phenomena in the human brain such diffusion on network [120, 130, 131, 145, 185], prion-like paradigm models [49, 172, 187, 188], and significant factors to predict disease progression [84, 85, 180].

2.3.1 Network diffusion model

In a series of papers, Raj and coworkers model the progression of neurodegenerative diseases by a diffusion process on the brain graph (connectome). The model is called the Network Diffusion Model (NDM). They claim that the model of misfolded protein, given by the graph Laplacian, predicts brain atrophy patterns of several neurodegenerative diseases such as AD and frontotemporal dementia [145]. The prediction of atrophies is based on the Laplacian eigenmodes (see Figure 2.3).

In NDM, dementia progression is modelled as a diffusion process on a brain network $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ whose nodes $v_i \in \mathcal{V}$ represent the ith cortical or subcortical grey matter structure, and whose edges, $e_{ij} \in \mathcal{E}$, represent white-matter fibre pathways connecting structures i and j. Structures v_i come from parcellation of brain MRI, and connection strength, c_{ij}, is measured by fibre tractography [12]. Suppose the disease factor at time t at each node in the network is represented by the vector $\mathbf{x}(t) = \{x(v,t), v \in \mathcal{V}\}$. The equation of the network diffusion model is given by

$$\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} = -\beta \mathbf{H}\mathbf{x}(t), \tag{2.1}$$

where $\mathbf{H}$ is the normalised graph Laplacian and β is the diffusion rate. Suppose $d_i = \sum_k c_{ik}$ is the sum of connection strength at node i, we have

$$H_{ij} = \begin{cases} \dfrac{-c_{ij}}{\sqrt{d_i d_j}}, & \text{for } c_{ij} \neq 0, \\ \dfrac{d_i}{\sqrt{d_i d_j}}, & \text{for } i = j, \\ 0, & \text{otherwise}, \end{cases} \tag{2.2}$$

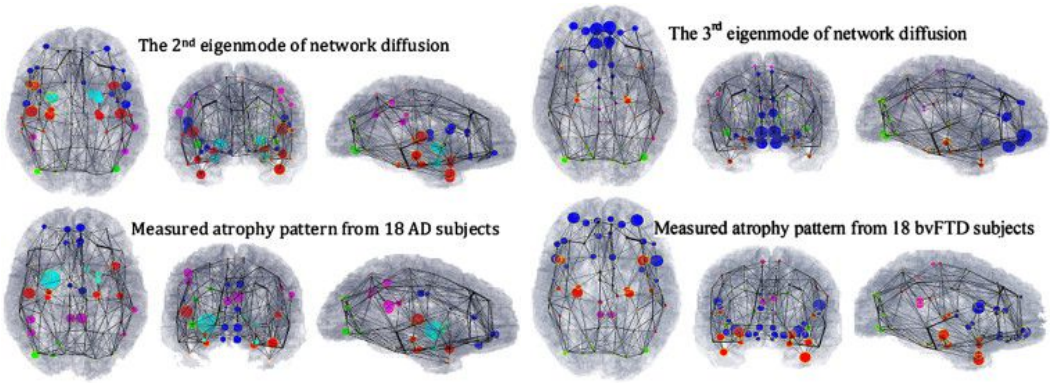
where each row and column of the Laplacian are normalised by their sums based on the modelling assumption that all brain regions are not the same size.

One feature of neurodegenerative diseases is the atrophy pattern of each disease due to the accumulation of toxic protein in the brain. Raj and coworkers hypothesised that cortical atrophy in region k due to accumulation is modelled as the integral of concentration

$$\phi_k(t) = \int_0^t x_k(s)\mathrm{d}s, \tag{2.3}$$

where atrophy in all modes increases with time and the slower rate induces more severe and widespread damage [145].

Further investigation on Laplacian eigenmodes indicated that global transmission inside the brain clusters the transport to areas such as superior and inferior areas [185]. Prediction of the atrophy can be improved when neuron polarity is considered [131]. These findings indicate that brain connectivity might explain the occurrence of neurodegeneration.



(a) Alzheimer's disease (b) Frontotemporal dementia

Figure 2.3: Visual correspondence between theoretical prediction and measured atrophy pattern. (a) Comparison for Alzheimer's disease and (b) Frontotemporal dementia. Predicted distribution was obtained from the Laplacian eigenmode (top) and measured the atrophy of 18 subjects on each disease(bottom). The size of the ball represents the amount of atrophy measured in the corresponding brain region. Blue, purple, green, red, and cyan colours represent frontal, parietal, occipital, temporal, and subcortical, respectively; adopted from Raj et al. [145].

The NDM model provides insight into revealing the transport of toxic proteins in the brain, however, we note a few plot holes. First, we raise a concern about the definition of toxic protein concentration. NDM uses the disease factor as its variable. However, whether the variable represents toxic protein concentration or any other measurement is unclear. When the model is compared against such concentration data, we need a clear definition of concentration. For example, we can define

a concentration as a biomarker abnormality percentage when we use toxic protein concentration from CSF measurement. We can also define a concentration as a tracer uptake ratio such as a standardised uptake value ratio (SUVR)[2] score obtained from PET scans. Second, we put a question to the assumption of using a normalised graph Laplacian. It usually comes from a random walk process instead of physical transport. Hence, NDM neglects physical properties such as mass conservation, in the absence of any explicit mass exchange terms such as production or clearance or specific reaction term, and Fick's law[3]. Third, the model also lacks non-linear interactions that occur from prion-like processes. In our understanding, it only focuses on how the misfolded proteins traverse in the brain. Thus, the purpose of this model to replicate real toxic protein propagation patterns remains unexplored [85].

2.3.2 Probabilistic model

A computational model incorporating the brain connectome, namely the epidemic spreading model (ESM) [85], is formulated as a probability rate of misfolded protein load in structural brain regions. The dynamic of the probability over time is determined by protein deposition and clearance probability. The protein deposition probability is modelled as the accumulation of external and internal infection of misfolded aggregates in the brain network.

The proposed model has stochastic features to describe the dynamic interactions between infectious misfolded protein and the brain's clearance mechanism. The system is governed by

$$\frac{\mathrm{d}P_i}{\mathrm{d}t} = (1 - P_i(t))\varepsilon_i(t) - \delta_i(t)P_i(t) + \mathcal{n}, \qquad \text{for } i = 1, ..., N, \tag{2.4}$$

where $P_i(t), \varepsilon_i(t), \delta_i(t)$ represent the temporal probability, the regional probability of misfolded protein, and the probability of being clean of misfolded protein at region i at time t, respectively. $\mathcal{n}$ denotes an additive noise due to possible stochastic processes, such as natural stochastic factors mediating misfolding protein aggregation [76] and is assumed to follow a Gaussian distribution with unknown mean μ and standard

[2]The standardised uptake value ratio is a medical imaging metric to quantify and compare the uptake radiotracers in different regions of the body, in this case the brain, from a PET scan. The intensity of the radiotracers that hit the target compound in the region of interest is recorded and standardised to a region that is not affected[31, 150, 151].

[3]Fick's law is a fundamental principle describing how substances move from high-concentration to low-concentration areas. In other words, there is no transport in the absence of a concentration gradient[144].

deviation σ, and

$$\varepsilon_i(t) = \sum_{j \neq i} Pa_{j \to i} \beta_j^{\text{ext}}(t - \tau_{ij}) + Pa_{i \to i} \beta_i^{\text{int}}(t) P_i(t), \tag{2.5}$$

$$\delta_i(t) = \delta_i(P_i, \delta_0) = 1 - e^{-\beta_0 P_i(t)}, \tag{2.6}$$

where $Pa_{j \to i}$ is the weighted anatomical connection probability between regions j and i. $\beta_j^{\text{ext}}(t - \tau_{ij}), \beta_i^{\text{int}}(t)$ become the extrinsic infection rate of region j at time $t - \tau_{ij}$ and the intrinsic infection rate of region i at time t given by

$$\beta_j^{\text{ext}}(t) = g(t)\beta_i(t), \tag{2.7}$$

$$\beta_i^{\text{int}}(t) = (1 - g(t))\beta_i(t), \tag{2.8}$$

$$\beta_i(t) = \beta_i(P_i, \beta_0) = 1 - e^{-\beta_0 P_i(t)}, \tag{2.9}$$

where τ_{ij} is the delay corresponding to the time the soluble misfolded protein departs from j and travels a connection distance $L_{j \to i}$. $\beta_0 \in [0, \infty)$ and $\delta_0 \in [0, \infty)$ are unknown constant parameters, with β_0 and δ_0 assumed to depend on the specific misfolded protein under study and the individual characteristics (e.g. genetic properties, lifestyle, environmental conditions).

Despite its complexity, the ESM provides insights into AD toxic protein propagation. The initial study using the ESM has demonstrated regional Aβ deposition characterisation and epicentre prediction by incorporating the structural brain connectome, protein clearance effect, and intercellular infection mechanism. This model is compared against Aβ datasets of healthy and diseased brains. It is suggested that a major reduction of Aβ clearance in early deposition results in excessive plaque formation. Further study has compared the ESM against τP PET data to investigate τP spreading hypothesis of Braak and Braak [180]. The model can explain about 70% of overall τP diffusion across the brain (Figure 2.4). However, the spreading behaviour cannot be captured accurately over time indicating that the model could not predict the further stages of AD pathology.

The ESM provides an insightful view of prion-like growth and spreading effect in the model, however, it has some disadvantages. We first note that parameters in the ESM lack a clear interpretation of functional changes associated with the disorder [187]. Its sophisticated structure does not directly use regional deposition but it highlights the importance of using the structural brain connectome in the model. Similar to the NDM of Raj [145], this model uses a symmetric brain connectivity matrix.

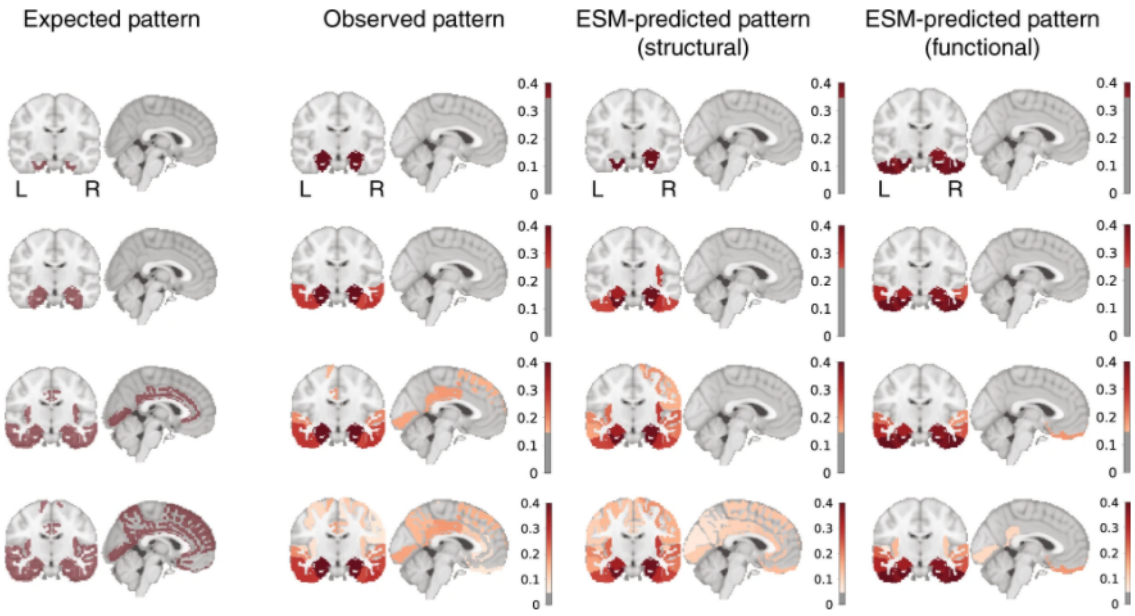


Figure 2.4: Validation of hypothesised, observed, and predicted pattern of τP spreading. The patterns of Braak initial stages (I-II) and late stages (V-VI) are given in the left column. The observed spreading patterns in τP PET imaging, ESM using structural connectome, and ESM using functional connectome from left to right. Warmer colours represent higher τP burden observed in the studied population; adopted from Vogel et al. [180].

2.3.3 Physics-based model

A physics-based model provides further mathematical modelling which accommodates reaction terms to describe misfolding proliferation. This non-linear interaction coupled with a diffusive process provides a more comprehensive perspective of prion-like propagation [187, 188]. A simple two-species model, known as the Heterodimer model, can be approximated by the well-known one-species Fisher-KPP model [187, 188]. Both the Heterodimer and Fisher-KPP models demonstrate biomarker abnormality of prion-like protein propagation (Figure 2.5). Let $\mathscr{P}$ and $\tilde{\mathscr{P}}$ represent healthy and misfolded protein molecules, respectively. The chain reaction between them is governed by

$$\begin{gathered} \mathscr{P} + \tilde{\mathscr{P}} \xrightarrow{k_{11}} \mathscr{P}\tilde{\mathscr{P}} \xrightarrow{k_{1'2'}} \tilde{\mathscr{P}}\tilde{\mathscr{P}} \xrightarrow{k_{2'2}} \tilde{\mathscr{P}} + \tilde{\mathscr{P}}, \\ \text{or} \\ \mathscr{P} + \tilde{\mathscr{P}} \xrightarrow{k_{12}} \tilde{\mathscr{P}} + \tilde{\mathscr{P}}. \end{gathered} \tag{2.10}$$

We can then formulate the reactions for the total amount of healthy protein $\mathscr{P}$ and misfolded protein $\tilde{\mathscr{P}}$. Let $p(\mathbf{r}, t)$ and $\tilde{p}(\mathbf{r}, t)$ represent the amount of healthy protein

$\mathscr{P}$ and the amount of misfolded protein $\tilde{\mathscr{P}}$ at location $\mathbf{r}$ and time t, respectively. The governing equation is given by

$$\frac{\partial p}{\partial t} = \mathrm{Div}(D_p \cdot \nabla p) + k_0 - k_1 p - k_{12} p \tilde{p}, \tag{2.11}$$

$$\frac{\partial \tilde{p}}{\partial t} = \mathrm{Div}(D_{\tilde{p}} \cdot \nabla \tilde{p}) \qquad - \tilde{k}_1 \tilde{p} + k_{12} p \tilde{p}, \tag{2.12}$$

where D_p and $D_{\tilde{p}}$ characterise the spreading of p and $\tilde{p}$, k_0 is the production rate of the healthy protein p, k_1 and $\tilde{k}_1$ are the clearance rates of p and $\tilde{p}$, and k_{12} is the conversion rate from the healthy to the misfolded conformation. It implies that in the healthy state, the equilibrium concentration of healthy protein is simply $p_0 = k_0/k_1$. Initially, the amount of healthy protein is much larger than the amount of misfolded protein, $p \gg \tilde{p}$, which implies that $\mathrm{d}p/\mathrm{d}t \approx 0$ and $\mathrm{Div}(D_p \cdot \nabla p) \approx 0$. With these assumptions, the equation provides an explicit estimate of the amount of healthy protein $p = k_0/(k_1 + k_{12}\tilde{p})$. We can now approximate the healthy protein concentration p using Taylor series evaluated at $k_{12}/k_1\tilde{p} = 0$, to obtain $p = k_0/k_1[1 - \tilde{p}k_{12}/k_1]$. Re-parameterising in terms of the misfolded protein concentration c, we now assume $0 \leq c \leq 1$ so that we have

$$\frac{\partial c}{\partial t} = \mathrm{Div}(D \cdot \nabla c) + \alpha c(1 - c) \tag{2.13}$$

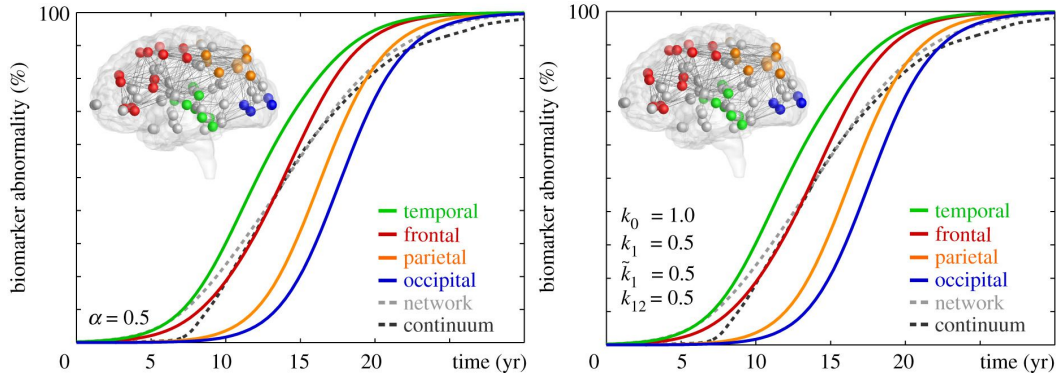
where $D = D_{\tilde{p}}$, $c = \tilde{p}/\tilde{p}^{\max}$, with $\tilde{p}^{\max} = \alpha k_1^2/k_{12}^2 k_0$ and $\alpha = k_{12}k_0/k_1 - \tilde{k}_1$. The spreading of misfolded protein can be either purely isotropic, such that $\mathrm{Div}(D \cdot \nabla c) = d\nabla c$, or anisotropic along pronounced communication networks within the brain. Importantly, in this model, the growth rate α of the misfolded protein concentration c increases linearly with the conversion rate from the healthy to the misfolded conformation k_{12} and with the equilibrium concentration of healthy protein $p_0 = k_0/k_1$ and decreases linearly with the clearance rate $\tilde{k}_1$.

These models can be coarse-grained by accommodating diffusion on a graph as the surrogate for the continuum Laplacian [49]. To model diffusion on a network, Weickenmeier and coworkers discretise (2.13) on an undirected graph where the concentration of misfolded proteins c_i at all $i = 1, ..., N$ nodes and express the change in the concentration as

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} = -\kappa \sum_{j=1}^{N} L_{ij} c_j + \alpha c_i (1 - c_i), \tag{2.14}$$

where κ characterises the global diffusion between regions and α is the net of local production and clearance of misfolded protein. A central element is the weighted

graph Laplacian L_{ij}, a square matrix, which we construct from the adjacency matrix A_{ij}. The graph Laplacian L_{ij} is the difference between the degree matrix and the adjacency matrix, $L_{ij} = d_i\delta_{ij} - A_{ij}$, where d_i denotes the degree of node i as stated in section 2.3.3. The graph Laplacian of the physics-based models is different from the proposed NDM model and incorporates physical processes. Furthermore, the Heterodimer model has been extended by considering Aβ and τP protein interaction showing that Aβ deposition can trigger the surge of secondary tauopathy [172]. Another extension has been proposed by considering different sizes of the protein molecule and more specific conformation processes, given by the Smoluchowski model [51, 49].



(a) Simulation using Fisher-KPP model (b) Simulation using Heterodimer model

Figure 2.5: Prion-like propagation simulations of Fisher-KPP and Heterodimer model. Biomarker abnormality of protein propagation across the individual brain lobes; adapted from Fornari et al. [49].

The physics-based models have essential advantages in developing an understanding of toxic protein transport in the brain through mathematical modelling. Despite the disease's complexity, this type of model captures three fundamental features of neurodegenerative diseases: toxic seeding, aggregate accumulation, prion-like proliferation and spreading throughout the brain. The model can summarise the regular sigmoid-like evolution of neurodegenerative biomarkers [49, 187, 188]. The reaction-diffusion of the Fisher-KPP is a minimal model that can predict the growth and spread of different neurodegenerative diseases. Simulations consistently suggest that anisotropic diffusion is a key ingredient revealing the underlying neurodegeneration and atrophy pattern formation mechanism. It is believed that mathematical models of neurodegeneration should incorporate networks of interconnected brain regions [49, 145]. Further observation using this model highlights the heterogeneity of brain networks and suggests some theoretical treatments. The topology of the brain

indicates the nature of AD's slow progression in the brain. Clinicians have used the Fisher-KPP model for further hypothesis testing of intracellular spreading by comparing the network model against τP data [150]. Using a weighted brain connectivity graph and brain regional distance, the study estimates the diffusion rate κ and the production rate α, and shows a good correlation with the observed τP PET signals. In general, physics-based models demonstrate the nature of neurodegenerative diseases.

We also note some drawbacks of the models. First, the intracellular τP spreading hypothesis is not fully confirmed due to a lack of longitudinal data. Second, despite developing more complex models incorporating various aggregate sizes, there is no data validation with Aβ or τP different aggregate sizes. Third, each model has a particular disadvantage. The Fisher-KPP model cannot fully demonstrate the interplay of biochemical and biomechanical degeneration for intermediate states. The initial heterodimer model cannot explain the effect of selective clearance in the brain. Fourth, the proposed models mostly focus on simulating global parameters and general propagation features. There is no explanation of the detailed spreading process on the brain network.

2.4 Objectives of the thesis

The main objective of my research is to understand the transport of neurodegenerative prion-like proteins within the brain networks from a mathematical modelling perspective. To some extent, this thesis is influenced by the physics-based mathematical model in [49] and [187] which reflects toxic protein propagation considering prion-like growth and spreading throughout the brain network. In this thesis, we ask particular questions about transport and network mathematical models. How does model selection affect toxic protein propagation on structural connectomes? What is the underlying mechanism of the systematic propagation of network-organised reaction-diffusion? What is the effect of asymmetry on toxic protein propagation on structural connectomes? To answer these questions, we address our research into three more specific problems: (1) staging problem, (2) arrival time problem, and (3) problem on systems with regional heterogeneity and directional connectivity.

2.4.1 Staging problem

The staging notion allows tracking the progression of toxic protein seeds and the associated damage. Motivated by Braak staging [18, 44], we define staging as an ordered sequence of brain regions of interest according to the time at which the average

toxic protein concentration reaches a given threshold. The staging problem consists in determining the sequence of brain regions of interest for different thresholds and model parameter values. A complete definition and discussion of the staging problem are addressed in Chapter 3. This work has been published in [144]. Our goal here is mainly to capture qualitative staging of τP protein as observed in Alzheimer's disease with specific questions:

1. What model correctly represents the spread of toxic protein along the axonal links? What parameter is crucial to toxic protein propagation, or to disease progression?

2. There are many methods for acquiring brain connectomes. Does it affect the resulting propagation pattern?

3. What is the effect of connectivity weight variation on the resulting pattern? Do seeds and the associated damage yield distinct patterns of propagation?

2.4.2 Arrival time problem

Focusing on toxic protein seed propagation, we consider the ordering sequence of brain regions of interest at a low concentration threshold. The time recorded at which the toxic protein concentration reaches a low threshold (equal to the initial value) is called *the arrival time*. The arrival-time problem consists in determining the arrival time at each node. This problem provides a meaningful insight to reveal systematic, perhaps hierarchical, invasion processes motivated by the study of toxic protein propagation in neurodegenerative diseases. This notion allows nodes associated with regions in the brain network to be ranked and reveals the staging of neurodegenerative disease. A thorough definition and discussion of the arrival-time problem are addressed in Chapter 4. This work has been published in [143].

Our specific goal for this problem is more than just to obtain a method to compute the arrival times since this problem can be solved numerically. Rather we want to obtain meaningful estimates and approximations of its dynamics. Therefore, our specific questions are:

1. How long does it take for the first seed to reach neighbouring nodes?

2. How does the first phase of the process depend on the parameters or topology of the system?

3. Once the initial invasion takes place, can we estimate the velocity of the invasion? How long does it take for the system to be fully invaded?

4. To obtain a decent analytical approximation of the solutions to understand other coupled processes.

2.4.3 Heterogeneous system and directed network problems

Evidence from different studies suggests that the transport process inside the brain can be influenced by many factors [21, 44, 63, 109]. Two possible factors are brain directed-connectivities [101, 165] and different brain regional properties (commonly called in the literature as heterogeneity) [37, 55, 122]. Directional bias is represented in the connectivity matrix by a directed network and regional heterogeneity is represented by different lower and upper concentration bounds between brain regions. The problems consist of determining dynamics from systems that have asymmetric features in the transport or varying kinetics of the model. Further definition and discussion of the problem are delivered in Chapter 5 for the heterogeneous system problem and Chapter 6 for the directed network system problem.

Our main goal is to understand the influence of particular factors such as directional bias and heterogeneity on the resulting dynamics in terms of arrival time and staging. Our specific questions are:

1. How long does a diseased seed invade the whole system in the presence of asymmetric links compared to its symmetric form?

2. If two networks have slight direction differences, do they show significantly distinct dynamics? How different are the dynamics of a perturbed system from its original system? Does the perturbation significantly affect the invasion process on a network?

3. What is the form of arrival time estimate derived for directed network systems?

4. How does directionality on structural connectomes change staging of τP propagation in the brain?

5. How does regional heterogeneity affect the staging of τP propagation in the brain?

6. What is the form of arrival time estimate derived for heterogeneous systems?

2.5 Contribution

The important findings of my research are:

1. Staging patterns can be identified using a network-organised Fisher-KPP model according to brain regions corresponding to the Braak stages [144]. This finding opens the possibility of individual-level assessment.

2. The underlying choice of brain connectome, the brain connectome's weight selection, and the model parameters' relative sizes could significantly influence the evolution of protein transport and tauopathy, and the three factors need to be considered concomitantly [144].

3. Arrival time estimate from the linear solutions obtained through the linearisation of the entire system of differential equations is universal and easily generalised to other systems [143].

4. The Lambert distance derived from the shortest path of edge arrival times on the associated network becomes an intrinsic property of the invasion process on a network. The method can provide a good initial guess for the linear arrival time. The approximation solution based on this estimate is an improvement from the linear solution [143].

5. The asymptotic solution based on the nonlinear variation of constants gives a uniformly valid approximation and has correct asymptotic properties. This method improves the solution given by the Lambert distance [143].

6. Both directed network and heterogeneous systems show that the Kernel of the defined graph Laplacian is not an equilibrium for the reaction-diffusion dynamics.

7. The linear method to estimate arrival time for both heterogeneous and directed network systems provides a good estimate and can capture the nontrivial features.

8. A similar notion to Lambert distance in [143] has been revisited for both heterogeneous and directed network systems. Large asymmetric features such that the parameter values are small strongly influence the accuracy of this method to estimate arrival times.

9. The asymptotic approach for both heterogeneous and directed network systems has been revisited. It becomes a better improvement to the linear and Lambert methods. The solution of this approach is given in a long expression in the form of a polylogarithm or hypergeometric function.

Chapter 3

Staging Problem in Neurodegenerative Disease

3.1 Overview

3.1.1 The hallmark of regional propagation

A hallmark of neurodegeneration disease is the regional spreading of toxic proteins in the brain. The protein aggregates are proposed to start in a particular brain region and then propagate along axonal links. Figure 3.1 illustrates the spreading of Aβ and τP in the AD brain. Post-mortem studies reveal that Aβ initially accumulates in a large area of the isocortex before invading the neocortex and cerebellar cortex. Meanwhile, τP is observed to accumulate in the entorhinal and invade other synaptically-connected brain regions. As the development of mathematical and computational models for neurodegeneration increases, different studies race to capture the signature pattern of toxic protein spreading in the brain.

In modelling neurodegenerative disease, a desirable property for any mathematical model is to reproduce the structured regional spreading and hierarchical patterns of progression called *staging*. The six-stage AD progression suggested by a seminal work in 1991 is a striking trait of the disease [18]. Mathematical models have been used to reproduce such staging patterns and data, e.g. PET scans [31]. A probabilistic model called ESM can accurately capture the region-by-region toxic protein propagation in lower stages. However, it failed for further stages when compared to scans [180]. Continuous models can also capture the lobe-to-lobe spreading in the sense of a more general propagation [49]. All models that can capture this trait agree that the entorhinal region in the brain is the seeding location to obtain the desired pattern despite their implication of model selections for an observed staging sequence.

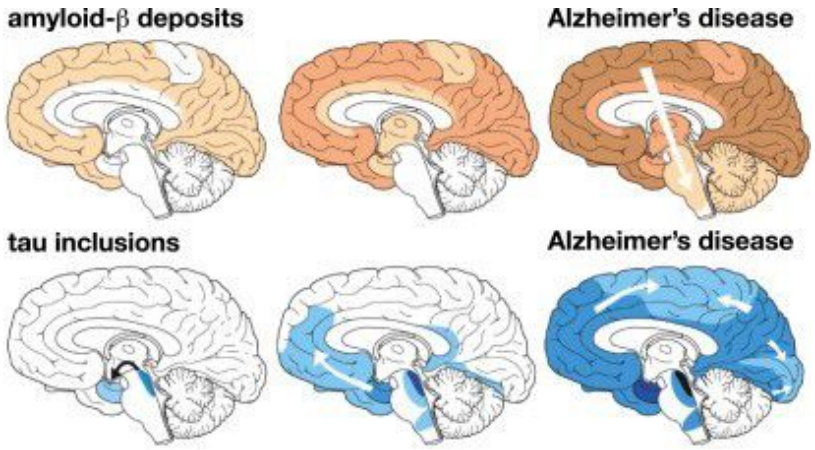


Figure 3.1: Toxic protein inclusion in AD. a. Amyloid-β aggregate inclusion in the diseased brain, b. τP aggregate inclusion in the diseased brain; adapted from Weickenmaier et al. [188]

In this chapter, we address questions related to model selection for computational studies of Alzheimer's disease through *the staging problem*. We will discuss the impact of model selection for prion-like mathematical models with respect to an observed staging sequence. As a case study, we will consider the propagation of toxic τP in AD. We compare results to patterns motivated by a recent study that uses a six-stage Braak pattern for τP in AD. Our study highlights the nuanced role of each choice in network neurodegeneration modelling. In particular, we show that choosing graph Laplacian weights, connectome resolution, and variation model parameters can significantly alter the landscape of observed staging patterns.

3.1.2 Computational disease staging

Disease staging is crucial for understanding disease progression and classifying illness severity. It is a method that defines discrete points to categorise increasing levels of severity according to a collection of observable clinical criteria. For instance, staging with three stages [62] classifies disease progression into stage 1 when conditions with no complications or problems with minimal severity, stage 2 when conditions with significantly increased risk of complications, and stage 3 when conditions with multiple site and systemic impairment, which may lead to death. This method clusters individuals into groups based on observed criteria. Staging is extremely meaningful for clinicians because it allows them to choose a therapeutic modality and evaluate the clinical significance of prognosis [36]. It has been widely used for many diseases including neurodegenerative diseases [18, 20] where the severity levels can be associated

with infected organs.

Scientists have used various computational methods to classify disease staging. In AD, a stage sequence is determined by ordering infected brain regions into a propagation order. There are two common methods of determining the sequence of infected brain regions according to biomarker levels. One can apply statistical analysis to empirical data to obtain the sequence. The other method is to apply a model and compute the average abnormality of different brain regions to replicate the signature propagation.

To define staging, one can use imaging data such as brain volumes from MRI scans or biomarker tracer levels from PET scans [31]. The approach is to process data according to grey and white matter segmentation. The SUVR score of various regions is classified into categories using the Z-score cutoff. The procedure determines the order of cortical spreading for participants with a Z-score above the cutoff and counts the number of participants categorised into the same group. This approach allows one to obtain the image-based τP stage as a useful biomarker for AD's clinical and pathological progression. However, this study is limited to the available data. For instance, the study uses cross-sectional data that only shows each patient's one-time medical record. This limitation hinders understanding the actual dynamics of spreading patterns.

3.1.3 Staging problem

The central problem discussed in this chapter is the *generalised staging problem*. We define *staging* as the spatiotemporal sequence of an invasion process on a network. Each element in the sequence is a disease stage associated with one or more regions in a network.

We consider a general spatiotemporal system with a form

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{f}(t, \mathbf{p}; \boldsymbol{\Theta}), \tag{3.1a}$$

$$\mathbf{p}(0) = \mathbf{p}_0, \tag{3.1b}$$

where $\mathbf{p}(t) = (p_1(t), ..., p_N(t))$ denotes a normalised and dimensionless quantity, for example, a normalised protein concentration $0 \leq p_i(t) \leq 1$, evolving on a connected and undirected network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a collection of N nodes $\mathcal{V}$, representing regions of interest, and edge set $\mathcal{E}$, representing connections between these regions. The quantity $p_i(t)$ corresponds to the observed concentration in node v_i and $p_i(0)$ is the initial concentration at that node. The quantity $\boldsymbol{\Theta}$ represents the parameters

of the system 3.1a. We further assume that the dynamics of the system is such that starting from an initial condition, where all concentrations are taken to vanish except in a few nodes, and the system will evolve asymptotically to a state where all concentrations reach their maximal value.

Let Ω_j for $j = 1, ..., J$ be a nonoverlapping collection of nodes, such that $\Omega_j \subseteq \mathcal{V}$ and $\Omega_j \cap \Omega_k =$ if $j \neq k$. Let $T \in [0, 1]$ be an arbitrary threshold value. As the concentration evolves according to (3.1), we define the average concentration P_j

$$P_j = \frac{1}{|\Omega_j|} \sum_{i \in \Omega_j} p_i, \quad j = 1, ..., J,$$

and record the time when P_j first reaches the threshold T, $P_j(t_j) = T$. We call these times the *arrival times*. This process produces an ordering of the regions Ω_j according to the ordered arrival times. The ordered sequence $\{\Omega_{j_1}, \Omega_{j_2}, ..., \Omega_{j_J}\}$ is called an *observed staging pattern*. The *generalised staging problem* here is to ascertain the scope of observable staging patterns subject to varying $\boldsymbol{\Theta}$. Ultimately, we can compare the observed staging patterns to one or more desirable staging patterns resulting from an empirical study.

3.2 Toxic protein propagation model

In this section, we discuss the modelling of prion-like protein propagation. We first consider the transport process across the brain connectivities and extend the modelling by including the pathological protein multiplication.

3.2.1 Diffusion on a network

The term *Laplacian matrix* connects different concepts in mathematics. In general, the Laplacian matrix is a matrix representation of a discrete Laplacian operator on a graph. The matrix approximates the negative-signed continuous Laplacian obtained from the operator discretisation. Enforcing the continuous Neumann boundary condition to the operator endows the matrix with a conserving mass property that prevents quantity loss during diffusion. In a graph setting, the Laplacian matrix is also called the *graph Laplacian*. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected connected graph with N nodes $\mathcal{V}$ and M edges $\mathcal{E}$. The matrix form of graph $\mathcal{G}$ is given by the adjacency $N \times N$ matrix $\mathbf{A} = [a_{ij}]$ with entries $a_{ii} = 0$ for all i and $a_{ij} = a_{ji} = 1$, for an unweighted graph, if and only if nodes i and j are connected. In the case of a weighted graph, the

entries are $a_{ij} = a_{ij} \geq 0$. The sum of entries along row i denotes the degree of node i, $d_i = \sum_{j=1}^{N} a_{ij}$. The *standard Laplacian matrix* for an undirected graph is given by

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \tag{3.2}$$

where $\mathbf{D}$ is the degree matrix, a diagonal matrix with entries $D_{ii} = d_i$ for $i = 1, ..., N$. The Laplacian matrix has properties: (1) $\mathbf{L}$ is symmetric, (2) $\mathbf{L}$ is semipositive definite, and (3) $\mathbf{L}$ has real eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_N$, with corresponding $v_1, ..., v_N$ as eigenvectors.

The graph Laplacian $\mathbf{L}$ describes a diffusive process on a network. Let us define $p_i = p_i(t)$ to be a concentration of a substance at node i at time t and $\mathbf{p} = (p_1, ..., p_N)$. Then, concentrations evolve in a diffusive process as

$$\begin{aligned} \frac{\mathrm{d}p_i}{\mathrm{d}t} &= -\rho L_{ij} p_j, \quad i = 1, ..., N, \\ \mathbf{p}(0) &= \mathbf{p}_0, \end{aligned} \tag{3.3}$$

where ρ denotes the diffusive rate. The concentrations at time t are then given by the exponential matrix

$$\mathbf{p}(t) = e^{-t\rho\mathbf{L}}\mathbf{p}_0,$$

and can be written as an expansion over the eigenvectors [136]

$$\mathbf{p}(t) = \sum_{i=1}^{N} k_i e^{-\rho t \lambda_i} \mathbf{v}_i, \tag{3.4}$$

where

$$k_i = \frac{\mathbf{v}_i \cdot \mathbf{p}_0}{\mathbf{v}_i \cdot \mathbf{v}_i}.$$

Since $\lambda_1 = 0$, the second smallest eigenvalue of $\mathbf{L}$, λ_2, is the dominant factor in concentration spreading and evolution.

Different studies have used different Laplacian matrices to study neurodegenerative and brain function. We note that studies of neurodegenerative disease in [1, 131, 145] and studies of the brain in [7, 8] have used normalised graph Laplacian, while a study in [130] used a random-walk graph Laplacian. Although these Laplacian matrices can reveal important brain features and neurodegeneration features, a natural question arises whether the standard, normalised, or other graph Laplacians are suitable for modelling neurodegenerative disease.

We note that modelling physical protein transport in the brain must satisfy some basic properties. Let us define a general graph Laplacian $\mathcal{L}$ where the concentration evolves according to

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathcal{L}\mathbf{p}, \tag{3.5}$$

without any sink or source nodes. We initially assume that each region of interest (ROI) has the same volume. This assumption leads to a mass conservation condition:

$$\sum_{i=1}^{n} \frac{\mathrm{d}p_i}{\mathrm{d}t} \equiv \mathbf{1} \cdot \frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = 0, \tag{3.6}$$

where $\mathbf{1} = (1, \ldots, 1)$ is the one vector. Using (3.5), this condition implies

$$\mathbf{1} \cdot \mathcal{L} \cdot \mathbf{p} = 0, \tag{3.7}$$

which must be true for all $\mathbf{p}$. Hence we have the condition

$$\text{(C1):} \quad \mathbf{1} \cdot \mathcal{L} = \mathbf{0}, \tag{3.8}$$

where $\mathbf{0} = (0, \ldots, 0)$ is the null vector. Fick's condition states that there is no transport in the absence of a concentration gradient. This statement is equivalent to the condition

$$\text{(C2):} \quad \mathcal{L} \cdot \mathbf{1} = \mathbf{0}. \tag{3.9}$$

When the graph Laplacian is symmetric, conditions (C1) and (C2) are mathematically equivalent. In addition to (C1) and (C2) we also consider a *robustness condition*,

$$\text{(C3): conditions (C1) and (C2) hold for all weighted adjacency matrices} \tag{3.10}$$

Condition (C3) guarantees, in particular, that variations in the weighted adjacency matrix should not incur a loss of the fundamental physical principles reflected by conditions (C1) and (C2). This condition is an important practical requirement because graph Laplacian weightings for structural connectomes may have different values depending on the number of fibres and their lengths. Although the dynamics of toxic proteins do not conserve mass due to aggregation or removal through clearance, the transport part of the model should preserve mass. Otherwise, it would require further assumptions to explain protein addition or removal because of the transport process between two nodes. We, therefore, insist on mass conservation by the diffusion part of the model, whereas other terms should model growth and clearance.

The two properties above must be addressed in modelling prion-like protein transport within the brain. It has consequences on the choice of the graph Laplacian. There are different ways of multiplying a standard graph Laplacian with a degree matrix to obtain a normalised graph Laplacian [1, 7, 8, 130, 131, 145]. To our knowledge, the sum of powers of degree matrices is never larger than 1. Therefore, we introduce the family of graph Laplacian

$$\mathbf{L}_{\alpha,\beta} = \mathbf{D}^{1-\alpha-\beta} - \mathbf{D}^{-\alpha}\mathbf{A}\mathbf{D}^{-\beta} = \mathbf{D}^{-\alpha}\mathbf{L}\mathbf{D}^{-\beta}, \tag{3.11}$$

with $\alpha, \beta \in [0,1]$ and $\alpha + \beta \leq 1$. The standard graph Laplacian corresponds to the choice $\alpha = \beta = 0$. That is, $\mathbf{L} = \mathbf{L}_{0,0}$ and we note that since the network is connected, the degree of each node is strictly positive and the inverse of $\mathbf{D}$ is well defined. We can now state the main result of this section

Proposition 1. *Suppose $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a connected undirected network with weighted adjacency matrix $\mathbf{A}$. Then, the standard graph Laplacian $\mathbf{L}_{0,0} = \mathbf{L}$ is the only member of the family* (3.11) *that simultaneously satisfies conditions (C1), (C2) and (C3).*

Proof. From the identities

$$\mathbf{D}\cdot\mathbf{1} = \mathbf{1}\cdot\mathbf{D} = \mathbf{A}\cdot\mathbf{1} = \mathbf{1}\cdot\mathbf{A} = \mathbf{d}, \tag{3.12}$$

where $\mathbf{d} = (d_1, ..., d_N)$, it follows that

$$\mathbf{L}\cdot\mathbf{1} = \mathbf{1}\cdot\mathbf{L} = \mathbf{0}. \tag{3.13}$$

Hence the standard Laplacian $\mathbf{L}$ satisfies conditions (C1) and (C2). Condition (C3) is satisfied because these relationships do not depend on the form of $\mathbf{A}$.

Next, we show that $\mathbf{L}_{0,0}$ is the only member of the family (3.11) satisfying all of (C1), (C2) and (C3). Let $\alpha, \beta \in [0,1]$, with $\alpha + \beta \leq 1$, be arbitrary but fixed with at least one of $\alpha > 0$ or $\beta > 0$. First, consider conditions (C2) and (C3). From (3.11), and (C2) we have

$$\mathbf{L}_{\alpha,\beta}\cdot\mathbf{1} = \mathbf{D}^{1-\alpha-\beta}\cdot\mathbf{1} - \mathbf{D}^{-\alpha}\mathbf{A}\mathbf{D}^{-\beta}\cdot\mathbf{1} = \mathbf{0}.$$

The i^{th} component of the above equation is

$$d_i^{1-\alpha-\beta} - \sum_{j=1}^{N}\frac{a_{ij}}{d_i^{\alpha}d_j^{\beta}} = \frac{d_i}{d_i^{\alpha}d_i^{\beta}} - \sum_{j=1}^{N}\frac{a_{ij}}{d_i^{\alpha}d_j^{\beta}} = 0.$$

Multiplying through by $d_i^{\boldsymbol{a}}$ and using the definition of $\mathbf{D}$ in terms of $\mathbf{A}$ yield

$$\sum_{j=1}^{N}\left(\frac{a_{ij}}{d_i^{\boldsymbol{b}}}-\frac{a_{ij}}{d_j^{\boldsymbol{b}}}\right)=\sum_{j=1}^{N}\frac{a_{ij}(d_j^{\boldsymbol{b}}-d_i^{\boldsymbol{b}})}{d_j^{\boldsymbol{b}}d_i^{\boldsymbol{b}}}=0. \tag{3.14}$$

Since this identity must be respected for all undirected connected graphs (condition (C3)), it must be independent of $\mathbf{A}$. Hence we must have

$$d_j^{\boldsymbol{b}}-d_i^{\boldsymbol{b}}=0, \tag{3.15}$$

for all pairs (i,j). This condition is satisfied by either $\boldsymbol{b}=0$ or $d_i=d_j$. If $d_i=d_j$ for all nodes and since there are at least three nodes in the network, one can change the weight of a single edge connected to one of the ROI, say i, but not the other by adding an arbitrarily small amount $0<\epsilon\ll 1$ to that weight, with the effect of changing d_i but not d_j. Hence, the equality $d_i=d_j$ cannot hold under the robustness assumption and we conclude that $\boldsymbol{b}=0$ is the only condition for which (C2) and (C3) hold simultaneously.

Next we consider the conditions (C1) and (C3). From (3.11), and (C1) we have

$$\mathbf{1}\cdot\mathbf{L}_{\boldsymbol{a},\boldsymbol{b}}=\mathbf{1}\cdot\mathbf{D}^{1-\boldsymbol{a}-\boldsymbol{b}}-\mathbf{1}\cdot\left(\mathbf{D}^{-\boldsymbol{a}}\mathbf{A}\mathbf{D}^{-\boldsymbol{b}}\right)=\mathbf{0}.$$

The j^{th} component of the column vector corresponding to the above equation states that

$$d_j^{1-\boldsymbol{a}-\boldsymbol{b}}-\sum_{i=1}^{N}\frac{a_{ij}}{d_j^{\boldsymbol{a}}d_i^{\boldsymbol{b}}}=\frac{d_j}{d_j^{\boldsymbol{a}}d_j^{\boldsymbol{b}}}-\sum_{i=1}^{N}\frac{a_{ij}}{d_i^{\boldsymbol{a}}d_j^{\boldsymbol{b}}}=0,$$

must hold identically. Multiplying through by the common term $d_j^{\boldsymbol{b}}$ and using the symmetry of the adjacency matrix yield

$$\sum_{i=1}^{N}\left(\frac{a_{ij}}{d_j^{\boldsymbol{a}}}-\frac{a_{ij}}{d_i^{\boldsymbol{a}}}\right)=\sum_{i=1}^{N}\frac{a_{ij}(d_i^{\boldsymbol{a}}-d_j^{\boldsymbol{a}})}{d_j^{\boldsymbol{a}}d_i^{\boldsymbol{a}}}=0. \tag{3.16}$$

As in the previous case, the robustness condition (C3) implies that

$$d_i^{\boldsymbol{a}}-d_j^{\boldsymbol{a}}=0, \tag{3.17}$$

for all pairs (i,j) and we conclude that mass conservation and robustness imply that $\boldsymbol{a}=0$. Conditions (C1), (C2) and (C3) concomitantly imply that $\boldsymbol{a}=\boldsymbol{b}=0$ must follow. □

In conclusion, diffusion on a network must only show concentration spreading across the connectome without self-loss or self-addition in the absence of other dynamical processes. In this section, we showed that the transport of toxic protein must follow mass conservation and Fick's condition which is satisfied by the standard graph Laplacian. We will use this type of graph Laplacian for the remaining of our work in this thesis.

3.2.2 Prion-like model on a network

Many mathematical models for neurodegeneration take advantage of the prion hypothesis, which asserts that the progression of proteopathy in AD follows a prion-like mechanism. The mechanism of toxic protein multiplication follows the heterodimeric process proposed by Prusiner depicted in Figure 3.2. There are two protein states: the healthy isoform $\mathscr{P}$ and the toxic isoform $\tilde{\mathscr{P}}$. The toxic protein can recruit healthy proteins and induce their toxic configuration. The new toxic protein chain is then fragmented and forms separated smaller molecules. The reactions can be collected and represented as a single kinetic represented with the rate k_{12} as follows,

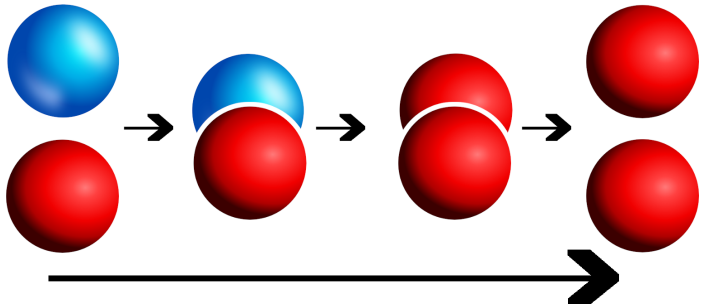


Figure 3.2: The reaction chain of the Heterodimer model. The blue molecule is the healthy isoform (namely $\mathscr{P}$) and the red molecule is the toxic isoform (namely $\tilde{\mathscr{P}}$. Two distinct isoforms combine and produce two toxic isoforms.

$$\mathscr{P} + \tilde{\mathscr{P}} \xrightarrow{k_{11}} \mathscr{P}\tilde{\mathscr{P}} \xrightarrow{k_{1'2'}} \tilde{\mathscr{P}}\tilde{\mathscr{P}} \xrightarrow{k_{2'2}} \tilde{\mathscr{P}} + \tilde{\mathscr{P}}, \tag{3.18a}$$

or

$$\mathscr{P} + \tilde{\mathscr{P}} \xrightarrow{k_{12}} \tilde{\mathscr{P}} + \tilde{\mathscr{P}}. \tag{3.18b}$$

The toxic protein can recruit healthy protein at a rate k_{11} and induce their toxic configuration at a rate $k_{1'2'}$. The new toxic protein chain is then fragmented at a rate $k_{2'2}$. Let $p = p(t)$ and $\tilde{p} = \tilde{p(t)}$ represent the concentration of healthy $\mathscr{P}$ and toxic

$\tilde{\mathscr{P}}$ proteins at time t, respectively. The chain reaction (3.18a) can be simplified into reaction (3.18b), and the Heterodimer model is given by,

$$\begin{aligned} \frac{\mathrm{d}p}{\mathrm{d}t} &= k_0 - k_1 p - k_{12} p\tilde{p}, \\ \frac{\mathrm{d}\tilde{p}}{\mathrm{d}t} &= \quad - \tilde{k}_1 \tilde{p} + k_{12} p\tilde{p}, \end{aligned} \tag{3.19}$$

where k_0 is the production rate of healthy protein, k_1 and $\tilde{k}_1$ are the clearance rates of healthy and toxic protein, and k_{12} is the conversion rate of healthy to toxic protein due to interaction of these two molecules.

There are two steady states of (3.19), namely the healthy state $p_0 = k_0/k_1$ and $\tilde{p}_0 = 0$, and the diseased state $p_\infty = \tilde{k}_1/k_{12}$ and $\tilde{p}_\infty = (k_0 k_{12} - k_1\tilde{k}_1)/(\tilde{k}_1 k_{12})$. The stability is discussed through the linearisation method based on the Jacobian matrix

$$\mathbf{J} = \begin{bmatrix} -k_{12}\tilde{p} - k_1 & -k_{12}p \\ k_{12}\tilde{p} & k_{12}p - \tilde{k}_1 \end{bmatrix}, \tag{3.20}$$

1. **Healthy state**. The eigenvalues λ evaluated at the stationary state $(p_0, \tilde{p}_0)$ are

$$\lambda_{0,1} = -k_1, \quad \lambda_{0,2} = \frac{k_0 k_{12} - k_1\tilde{k}_1}{k_1}, \tag{3.21}$$

   which shows that the local stability depends on the second eigenvalue. The healthy state is asymptotically stable when $k_0 k_{12}/k_1\tilde{k}_1 < 1$.

2. **Diseased state**. This stationary state shows that toxic protein may exist in the brain and is influenced by the toxic protein clearance rate and the infection rate. The eigenvalues λ evaluated at the stationary state $(p_\infty, \tilde{p}_\infty)$ are

$$\begin{aligned} \lambda_{\infty,1} &= -\frac{1}{2\tilde{k}_1}\left[k_0 k_{12} + \mathscr{K}^{1/2}\right], \\ \lambda_{\infty,2} &= -\frac{1}{2\tilde{k}_1}\left[k_0 k_{12} - \mathscr{K}^{1/2}\right], \end{aligned} \tag{3.22}$$

   where $\mathscr{K} = (k_0 k_{12})^2 - 4\tilde{k}_1^2\left(k_0 k_{12} - k_1\tilde{k}_1\right)$. This steady state exists when $k_0 k_{12}/k_1\tilde{k}_1 > 1$ and is asymptotically stable when it exists.

We now model a reaction-diffusion process that describes the protein multiplication and spreading throughout a network. We use the heterodimer coupled with the diffusion on a network. We introduce the concentration of healthy and toxic protein

at node i, p_i and $\tilde{p}_i$, and their transport across the network is characterised by the standard Laplacian $\mathbf{L}$

$$\begin{aligned}\frac{\mathrm{d}p_i}{\mathrm{d}t} &= -\rho\sum_{j=1}^{N} L_{ij}p_i + k_0 - k_1 p_i - k_{12}p_i\tilde{p}_i,\\ \frac{\mathrm{d}\tilde{p}_i}{\mathrm{d}t} &= -\rho\sum_{j=1}^{N} L_{ij}\tilde{p}_i \quad - \tilde{k}_1\tilde{p}_i + k_{12}p_i\tilde{p}_i,\end{aligned} \tag{3.23}$$

where $i = 1, ..., N$.

Close to the initial healthy state, $p_i \gg \tilde{p}_i$, the system (3.23) can be approximated by a Fisher-KPP equation. The condition of $p_i \gg \tilde{p}_i$ and its close-to-uniform distribution across the network imply that $\mathrm{d}p_i/\mathrm{d}t \approx 0$ and $\sum_j L_{ij}p_j \approx 0$. With these assumptions, the amount of healthy protein in each node can be approximated by

$$k_0 - k_1 p_i - k_{12}p_i\tilde{p}_i = 0, \text{ thus } p_i = \frac{k_0}{k_1 + k_{12}\tilde{p}_i}, \tag{3.24}$$

which can be simplified using Taylor approximation to obtain an expression of toxic protein

$$\frac{\mathrm{d}\tilde{p}_i}{\mathrm{d}t} = -\rho\sum_{j=1}^{N} L_{ij}\tilde{p}_j + \left[\frac{k_{12}k_0}{k_1} - \tilde{k}_1\right]\tilde{p}_i - \frac{k_{12}^2 k_0}{k_1^2}\tilde{p}_i^2. \tag{3.25}$$

We define $c_i = \tilde{p}_i/\tilde{p}_{max}$ to obtain a Fisher-KPP expression of toxic protein concentration in a network,

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} = -\rho\sum_{j=1}^{N} L_{ij}c_j + \alpha c_i\left(1 - c_i\right), \tag{3.26}$$

where

$$\tilde{p}_{max} = \frac{\alpha k_1^2}{k_{12}^2 k_0}, \alpha = \frac{k_{12}k_0}{k_1} - \tilde{k}_1. \tag{3.27}$$

In the absence of diffusion, a similar behaviour to the Heterodimer model is observed for the homogeneous system of (3.26) where, in the initial healthy state, the toxic protein concentration is $c_0 = 0$, and evolves towards the diseased state given by the concentration $c_\infty = 1$. This uniform solution is the kernel of the Laplacian matrix $\mathbf{L}$. We will now show that the uniform diseased state of the system (3.26) is asymptotically stable when the local fixed point is stable.

Let $F(c) = \alpha c(1 - c)$ be the local vector field on an individual node. Let c_∞ be a linearly stable fixed point of

$$\frac{\mathrm{d}c}{\mathrm{d}t} = F(c). \tag{3.28}$$

Since the non-linear term F is the same at every node, we can extend this solution to the entire nodes by defining a vector $\mathbf{c}^* \in \mathbb{R}^N$ such that $\mathbf{c}_i^* = c_\infty$, $i = 1, ..., N$. The vector will be the fixed point for system (3.26) since $\mathbf{L} \cdot \mathbf{c}^* = \mathbf{0}$. We want to show that $\mathbf{c}^*$ is stable. The Jacobian matrix for (3.28) is

$$\mathcal{J} = F_c \mathbf{I} - \rho \mathbf{L}, \tag{3.29}$$

where $F_c = \mathrm{d}F/\mathrm{d}c(c_\infty) < 0$. Let λ_i and v_i be a pair of eigenvalue and associated eigenvector of graph Laplacian $\mathbf{L}$ for $i = 1, .., N$. The eigenvectors of $\mathbf{L}$ are also the eigenvectors of $\mathcal{J}$ with corresponding eigenvalues $F_c - \rho\lambda_i$, $i = 1, ..., N$.

$$\mathcal{J} v_i = [F_c \mathbf{I} - \rho \mathbf{L}] \cdot v_i = (F_c - \rho\lambda_i) v_i. \tag{3.30}$$

The eigenvalues, λ_i, are non-negative for $i = 1, ..., N$ and $F_c < 0$. Hence, the eigenvalues of system (3.26) around the fixed point vector $\mathbf{c}^*$, $\Lambda_i = F_c - \rho\lambda_i < 0$ (given that ρ is positive) for $i = 1, ..., N$. Hence, the uniform diseased state $\mathbf{c}^*$ is asymptotically stable. □

The evolution of the degeneration process in AD can be measured from the concentration of toxic proteins identified as biological markers [87]. Here, we introduce a seed protein biomarker to analyse staging in the Fisher-KPP network model. In clinical imaging data, the available staging patterns do not necessarily have a resolution at the level of a single connectome node. We then consider J regions $\Omega_1, \ldots, \Omega_J$. In each of these regions, we define

$$C_\Omega(t) = \frac{1}{c_\infty |\Omega|} \sum_{i=1}^{|\Omega|} c_i(t), \tag{3.31}$$

where $|\Omega|$ is the number of nodes in Ω.

The equation (3.26) governs the evolution of toxic proteins that captures the fundamental characteristic of AD, i.e., the pathogenic protein's growth and spread. We will now focus on the Fisher-KPP model to explain the progression of toxic proteins instead of the Heterodimer model.

3.2.3 Damage model

We have proposed a Fisher-KPP model and biomarker abnormality measure to understand how toxic proteins propagate across the brain network. However, the model is limited to seed protein concentration and does not tell how the accumulating toxic

protein affects the brain over time. We propose a new measure q_i denoting the damage at node i ($0 \leq q_i \leq 1$) coupled with (3.26),

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = \delta c_i(1 - q_i), \quad q_i(0) = 0, \tag{3.32}$$

where $i = 1, ..., N$ and δ is the overall pathology progression rate constant. In the absence of non-linear interaction ($\alpha = 0$), we obtain the diffusion system coupled to damage.

The pathology burden of protein deposition is defined as a discrete sum of damage across nodes inside Ω, similar to toxic seed biomarker abnormality. We then define a damage measure on each region $\Omega_1, ..., \Omega_J$ given by

$$Q_\Omega(t) = \frac{1}{|\Omega|} \sum_{i=1}^{|\Omega|} q_i(t), \tag{3.33}$$

where $|\Omega|$ is the number of nodes in Ω.

3.3 Implementation and results

We apply our approach to a six-stage Braak pattern of τP in AD motivated by an observation that seed-competent tau precedes tau aggregation. A study in [44] has advanced the notion that τP seeds precede NFT pathology in a similar structured manner observed in the literature that refers to the evolution of τP NFT or of SUVR levels. We use a model that tracks the progression of both τP seeds and τP NFT through equation (3.26) and (3.32), respectively. We simulate our model using a nondimensional form with fewer parameters. We rescale the model with $t \sim 1/\alpha$ and the system is now governed by

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} = -\beta \sum_{j=1}^{N} L_{ij} c_j + c_i \left(1 - c_i\right), \quad i = 1, ..., N, \tag{3.34a}$$

$$\frac{\mathrm{d}q_i}{\mathrm{d}t} = \sigma c_i(1 - q_i), \quad i = 1, ..., N, \tag{3.34b}$$

where $\beta = \rho/\alpha$ denotes the ratio between the diffusion and growth rates while $\sigma = \delta/\alpha$ denotes the ratio between the damage and growth rates. The initial condition is $c_s(0) > 0$ for seeding nodes s, otherwise $c_i(0) = 0$, while $q_i(0) = 0$ for all i. We will show how various aspects of model selection can strongly influence the observed regional progressions. We adopt the staging regions given in [44] and the same method can be generalised to the study of other staging regions given in [31, 41, 151, 180].

3.3.1 Braid diagram and braid surface

We begin with a braid diagram. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a fixed network and suppose that Ω_1, Ω_2, ..., Ω_J are a set of non-overlapping node regions. Suppose that T_1, T_2, ..., T_N are (biomarker) threshold values in the unit interval $[0, 1]$. A *braid diagram* depicts the index of the regions Ω_j as an abscissa and the threshold values T_k as an ordinate. As a dynamical system, such as (3.26), evolves on $\mathcal{G}$, we record the time $t_{j,k}$ at which a region, Ω_j, first achieves the threshold, T_k. We fix $t_{j,k} = \infty$ when a particular region never reaches a given threshold. We determine an ordering of the regions Ω_j for a fixed threshold index from the collection of recorded times. We visualise the ordering as a braid diagram after we put them together into a plot.

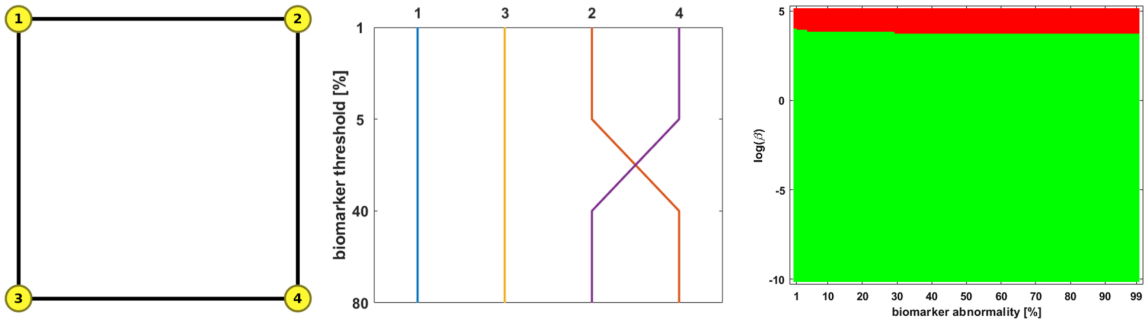
Staging outcomes demonstrate different patterns as we vary the system parameters. Our task is to assign a colour for each possible staging. We introduce a *braid surface* where that information summarises generalised braid diagrams. We plot them to a two-dimensional plot in a pair of one parameter value and one biomarker threshold corresponding to a staging colour. We can generate this surface by a simple algorithm. First, we discretise the values of a parameter of interest, such as β in (3.34a). Next, the system is solved for each discretised parameter. We obtain the observed staging pattern for each threshold value. We add any newly observed staging pattern to the list if it has not yet been encountered. At the end of the process, every pairing of discrete parameters and the threshold has been assigned to an observed staging pattern. We assign colours to the set of observed staging patterns and plot the colours to visualise the braid surface.

We give an example of a braid diagram in Figure 3.3. The graph consists of four nodes and the regions are $\Omega_i = \{i\}$ for $i = 1, 2, 3, 4$. The threshold values are $T_1 = 1\%$, $T_2 = 5\%$, $T_3 = 40\%$ and $T_4 = 80\%$. We determine staging from solving (3.34a) with $\ln(\beta) = 3.897$ and a synthetic weight matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 3.125 \times 10^{-7} & 5 \times 10^{-6} & 0 \\ 3.125 \times 10^{-7} & 0 & 0 & 1 \times 10^{-5} \\ 5 \times 10^{-6} & 0 & 0 & 1.5 \times 10^{-4} \\ 0 & 1 \times 10^{-5} & 1.5 \times 10^{-4} & 0 \end{bmatrix}.$$

This example shows that the regions achieve the threshold $T = 5\%$ in the order $\Omega_1 \to \Omega_3 \to \Omega_2 \to \Omega_4$. For the threshold $T = 40\%$, the observed ordering of regions changes to $\Omega_1 \to \Omega_3 \to \Omega_4 \to \Omega_2$. We can express these two observed staging patterns as the abbreviated notation. The green in the braid surface corresponds to the 1-3-2-4 staging sequence, while the red corresponds to the 1-3-4-2 staging sequence. A braid

diagram is useful for considering observed staging for a fixed set of model parameters; for instance, for a fixed value of β in (3.34a).



A A network with 4 regions of interest

B A 4-Threshold Braid Diagram

C A Braid Surface of Network **A**

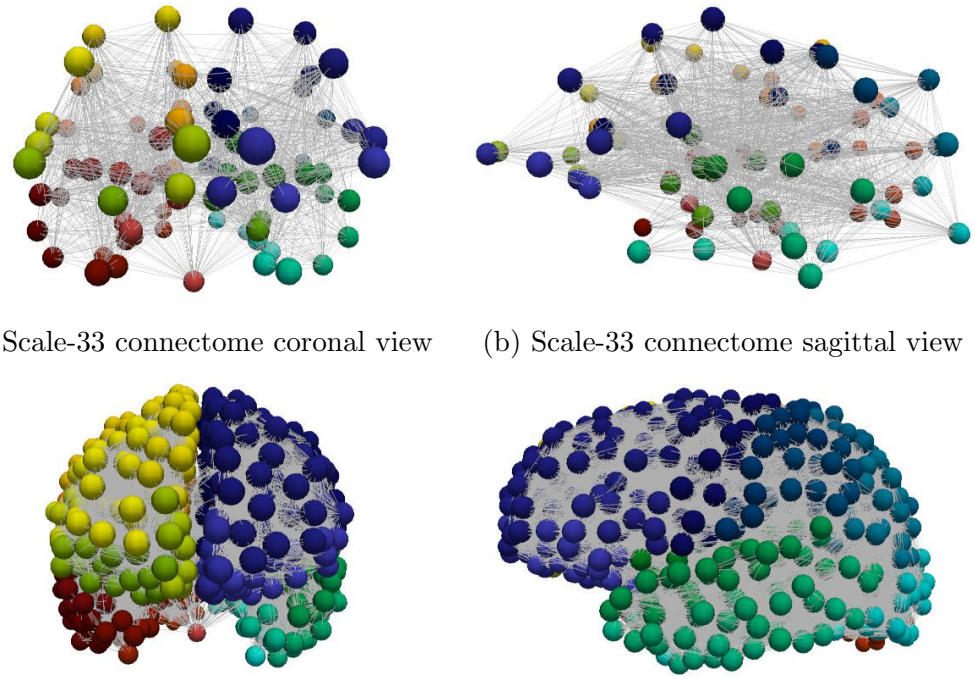
Figure 3.3: A network with four regions of interest and equal edge weight (left). A braid diagram (middle) is generated for a fixed value of β. A braid surface (right) showing the τP seed staging dynamics for the simple network as (β) varies. Two distinct staging patterns are produced for the network considered.

3.3.2 Graph weighting

The connectome is a discrete representation of the human brain given by a collection of nodes and edges representing parcellated brain areas and interactions between them. The functional and structural connectomes are networks commonly used to study brain function or neurodegenerative disease progression. Both graphs are obtained through MRI and diffusion tensor imaging (DTI) image processing, but they show distinct parcellation maps due to different purposes and regions acquired. The functional network is obtained by recording the signal in a resting state or while doing a particular task. It is used to study activities or functions within the brain [7, 8]. The structural connectome is the anatomical representation of the brain and is obtained by imaging the volume of the brain tissue. In this study, we will focus on structural connectomes.

We use a collection of individual patient data in a GraphML format downloaded from BrainGraph.org [102]. All of the structural connectomes considered in this thesis are constructed using the Lausanne multi-resolution atlas parcellation [40] with five levels of potential resolution. Each patient file accommodates the region nodes (with their name and coordinate) and interaction links between connected regions (with the length and fibre number). The edges, at all scales, include information regarding the

number of fibres (n_{ij}) and fibre length (ℓ_{ij}) associated with an edge connecting region i to region j. In the smallest resolution, a brain region is represented by one node, while the other resolutions may have more than one node for a region. Specifically, these distinct resolutions provide five different node scales, i.e. 83, 129, 234, 463 and 1015 nodes. Figure 3.4 shows the lowest and highest resolution connectome. The connectomes are constructed using deterministic and probabilistic methods. The deterministic connectomes are constructed using MRtrix [102, 175] and a deterministic streamlining with 20,000 streamlines and randomised seeding.



(a) Scale-33 connectome coronal view (b) Scale-33 connectome sagittal view

(c) Scale-500 connectome coronal view (d) Scale-500 connectome sagittal view

Figure 3.4: The structural connectome. The lowest resolution (a, b) with 83 nodes and the highest resolution (c, d) with 1,015 nodes.

In this study, we construct the probabilistic connectomes to assess our observations using the deterministic connectomes. We use the Connectome Mapping Toolkit in [40] to parcellate a high-resolution MNI reference template and the FSL [159, 189, 91] PROBTRACKX algorithm to provide the probabilistic tractography with 10,000 streamlines per voxel. We observe that the resulting connectomes are quite dense (approximately 7-12% sparsity). In this case, we are interested to study the effect of thresholding. We apply five thresholding techniques for our comparative analysis. We summarised the techniques in Table 3.1. The corresponding citation provides a description of the method. The last technique, called naive thresholding method,

Connectivity thresholding	Abbreviation	Citation
Disparity Filtering	DF	[155]
Doubly stochastic DF	DS	[157]
High salience skeleton	HSS	[64]
Noise-corrected backboning	NC	[38]
Naive cutoff	NV	N/A

Table 3.1: Thresholding methods for probabilistic connectomes

removes edge e_{ij} if the corresponding connectivity coefficient (n_{ij}) is below a given threshold value.

The diffusion on a network depends on some edge weights a_{ij}. Based on the available data, the connection between nodes in the brain is represented by an idealised uniform cylinder constructed from n_{ij} fibres and its length l_{ij}. We assume that all material properties are uniform. There are three plausible choices of weight that we may consider.

1. **Length-free transport**
   We assume that protein transport in the brain does not depend on the length of fibres but is limited to other instantaneous mechanisms such as cell-to-cell transport. This assumption implies that the number of fibres could be a suitable choice based on this assumption. Hence, we have $a_{ij} = n_{ij}$. The length-free weight is used in [1, 145].

2. **Ballistic transport**
   Another possible assumption is that fibre length affects the transport process. The weight is then penalised by the length indicating that a shorter fibre transports protein faster. Thus, we have $a_{ij} = n_{ij}/l_{ij}$. Assuming the weights of adjacency matrix $\mathbf{A}$ follow this choice, the overall rate ρ then has a dimension of velocity (length per time). It is reminiscent of front propagation given by a nonlinear reaction-diffusion process. The ballistic weight is used in [49, 51].

3. **Diffusive transport**
   The last assumption is that the transport is penalised by the length squared, where $a_{ij} = n_{ij}/l_{ij}^2$ when node i and j are connected. This choice is the discretisation of a continuous Laplacian over a regular grid and ρ has the dimension (length squared per time) of a diffusion constant. This choice is natural in the absence of other transport assumptions and is used in [172].

3.3.3 Language for staging

Using the braid surfaces, we study the staging problem for the model (3.34) of tauopathy in AD. We use five structural connectome ROIs to match the six regions used in [44]. Table 3.2 describes the ROIs used in our study.

Stage	Connectome ROI	Stage	Connectome ROI
Region I	entorhinal cortex	Region II	hippocampus
Region III	parahippocampal gyrus	Region IV	rostral and caudal anterior cingulate
Region V	cuneus, pericalcarine cortex, lateral occipital cortex, lingual gyrus		

Table 3.2: Connectome regions used in the staging problem.

Table 3.3 describes all observed regional staging patterns in (3.34). We determine each staging progression with Roman numerals, for instance, the I→II notation, at a concentration threshold $0\% < T \leq 100\%$. According to [18] of τP Braak staging, we

All observed computational staging patterns				
Suggested patterns Methods		$I \to II \to III \to IV \to V^*$		$I \to III \to II \to IV \to V^\dagger$
		$I \to II \to III \to V \to IV^\ddagger$		$I \to III \to II \to V \to IV^\ddagger$
Other patterns with entorhinal cortex origin		$I \to III \to V \to II \to IV$		$I \to III \to V \to IV \to II$
		$I \to III \to IV \to V \to II$		$I \to III \to IV \to II \to V$
		$I \to IV \to II \to III \to V$		$I \to IV \to III \to II \to V$
		$I \to IV \to V \to III \to II$		$I \to IV \to III \to V \to II$
		$I \to V \to III \to II \to IV$		
Patterns with extraentorhinal origin		Beginning with stage II		Beginning with stage III
		Beginning with stage IV		Beginning with stage V

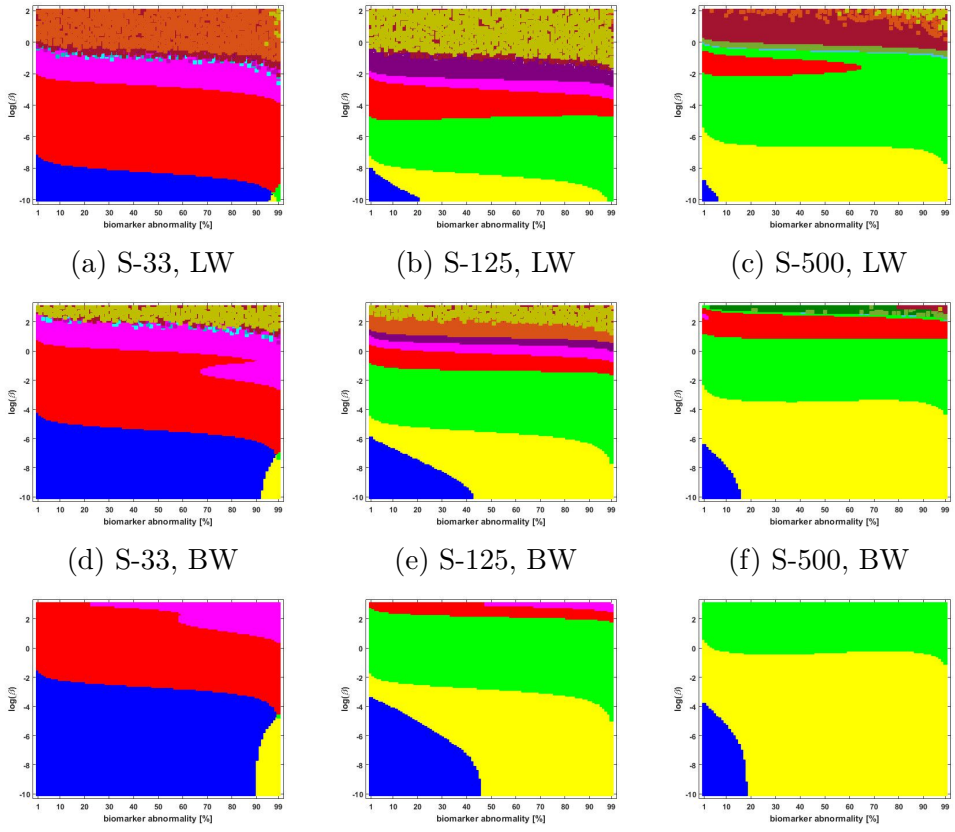
Table 3.3: Color coding for braid surfaces. * Progressive computational Braak staging. † Computational staging suggested by SUVR data. ‡ Additional potential computational staging, τP seeding.

note that $I \to II \to III \to IV \to V$ progression is an expected candidate of interest. However, we also find three progression sequences that may be of interest for both τP seeds and NFT when we explore the ADNI data. We call the set of identified staging patterns of potential interest *suggested patterns* and provide other patterns in Table 3.3.

3.3.4 Staging result using deterministic connectomes

We explore Braak staging pattern using deterministic connectomes by considering 1% to 99% biomarker abnormality level with 1% step size. We show the observed

staging for both τP seeds (Figure 3.5) and NFT (Figure 3.6). We present these results for Scale-33, Scale-125, and Scale-500 resolution connectomes using all three graph Laplacian weighting schemes. All figures showcase results for the model parameter value $\beta = \rho/\alpha$ in the range $-10 \leq \ln(\beta) \leq 2$.



(a) S-33, LW (b) S-125, LW (c) S-500, LW

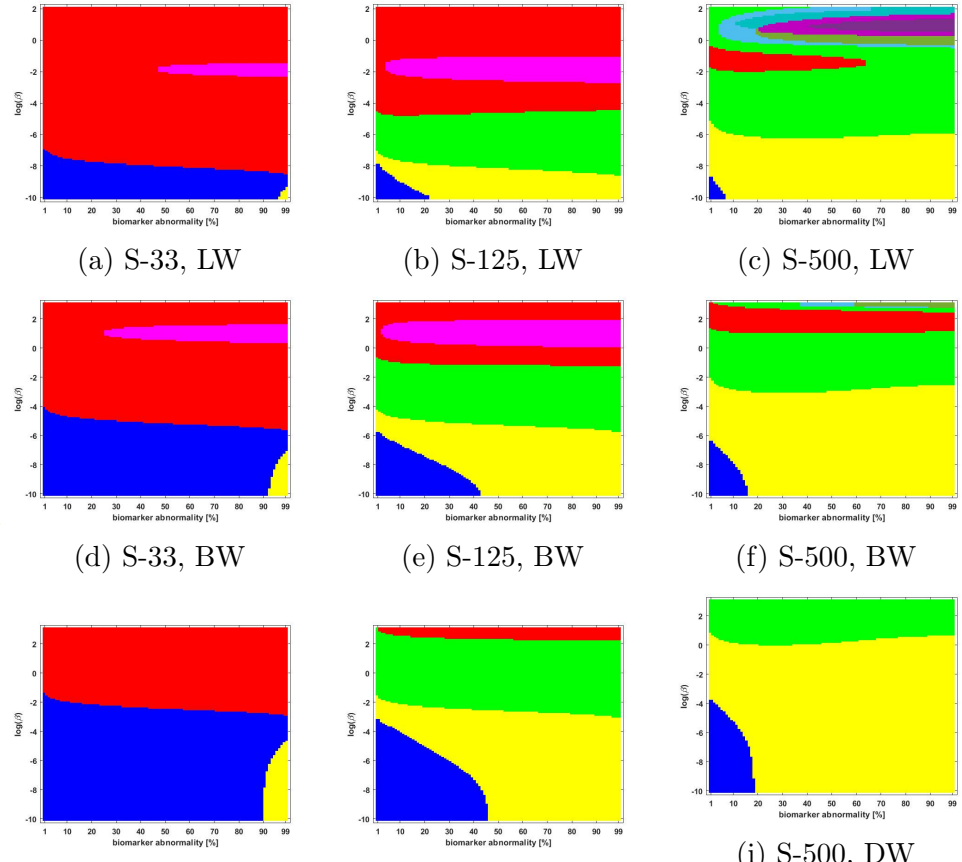
(d) S-33, BW (e) S-125, BW (f) S-500, BW

(g) S-33, DW (h) S-125, DW (i) S-500, DW

Figure 3.5: Observed computational (deterministic) connectome τP seed staging; length-free (top), ballistic (middle) and diffusive (bottom) weighting schemes. The x-axis determines the biomarker abnormality threshold $1\% < T \leq 100\%$ and the y-axis corresponds to $-10 \leq \ln(\beta) \leq 2$ for the parameter β in (3.34a).

3.3.5 Staging result using probabilistic connectomes

We explore Braak staging pattern using probabilistic connectomes by considering 1% to 99% biomarker abnormality level with 1% step size. We present these results for Scale-33, Scale-125, and Scale-500 resolution connectomes using all three graph Laplacian weighting schemes. All figures showcase results for the model parameter value β in the range $-10 \leq \ln(\beta) \leq 2$. In the case of probabilistic connetome, we explore the effects of the parameter β and thresholding methods on staging pattern in



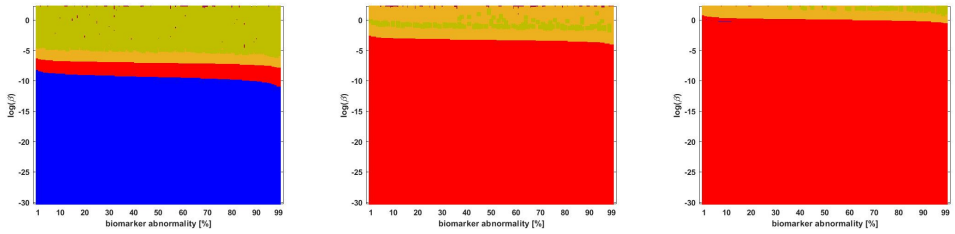
(a) S-33, LW (b) S-125, LW (c) S-500, LW

(d) S-33, BW (e) S-125, BW (f) S-500, BW

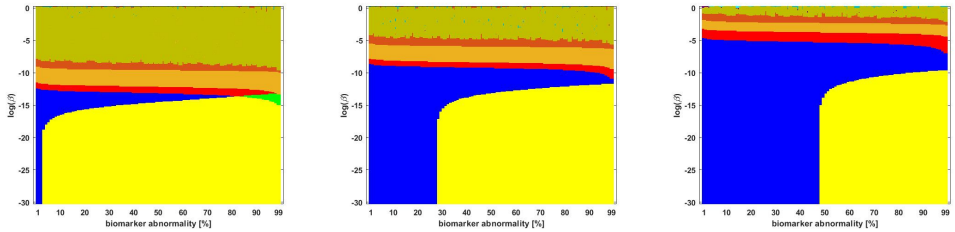
(g) S-33, DW (h) S-125, DW (i) S-500, DW

Figure 3.6: Observed computational (deterministic) connectome τP NFT staging with $\sigma = 1$ in (3.34b). Figure order, axis labels and axis ranges are identical to those of Figure 3.5.

(3.34). As an example, we present only one thresholding value for each thresholding method.

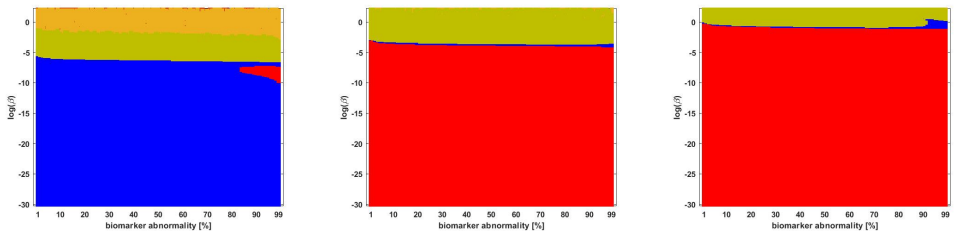


(a) S-33, LW (b) S-33, BW (c) S-33, DW

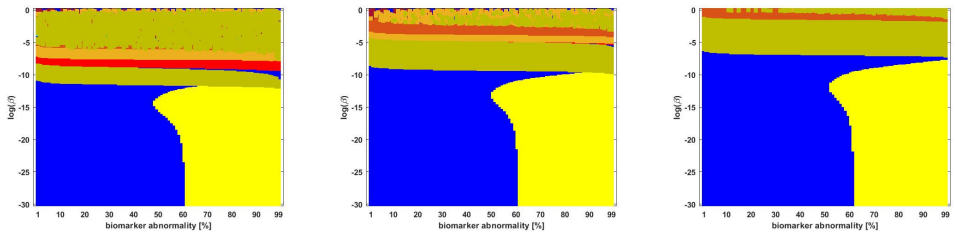


(d) S-500, LW (e) S-500, BW (f) S-500, DW

Figure 3.7: Observed computational (probabilistic) connectome τP seed staging. Density filter (DF) method at a threshold of 2×10^{-1}. The x-axis determines the biomarker abnormality threshold $1\% \leq T \leq 100\%$ and the y-axis corresponds to $-30 \leq \ln(\beta) \leq 0$ for the parameter β in (3.34b).

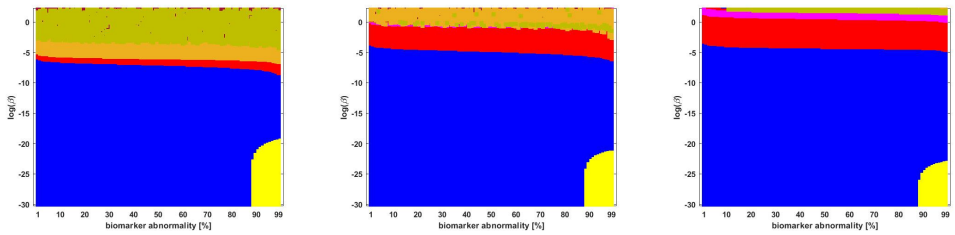


(a) S-33, LW (b) S-33, BW (c) S-33, DW

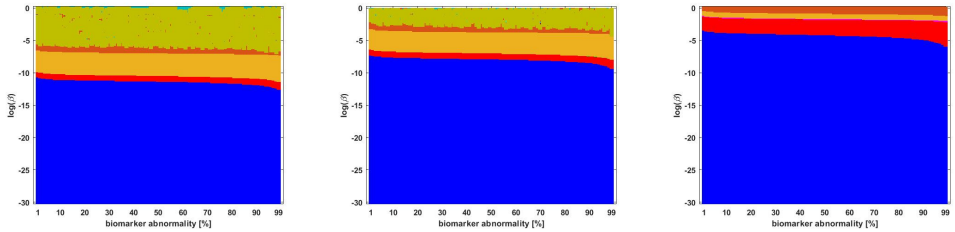


(d) S-500, LW (e) S-500, BW (f) S-500, DW

Figure 3.8: Observed computational (probabilistic) connectome τP seed staging. Doubly stochastic method at a threshold of 1×10^{-2}. Axes coincide with those of Figure 3.7

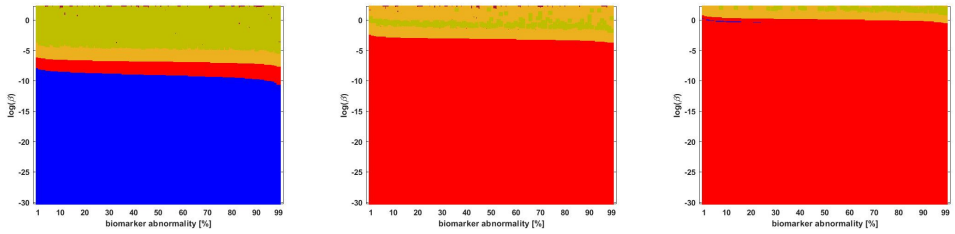


(a) S-33, LW (b) S-33, BW (c) S-33, DW

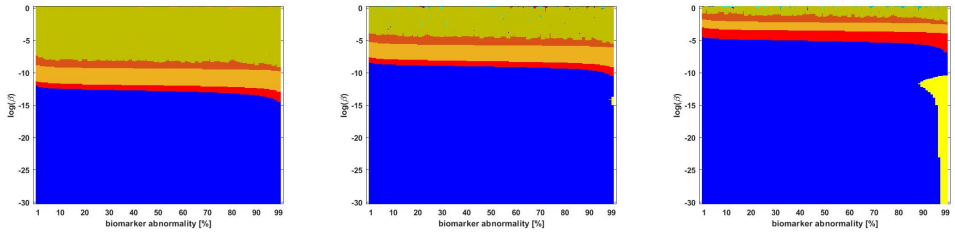


(d) S-500, LW (e) S-500, BW (f) S-500, DW

Figure 3.9: Observed computational (probabilistic) connectome τP seed staging. High salience skeleton at a threshold of 1×10^{-1}. Axes coincide with those of Figure 3.7

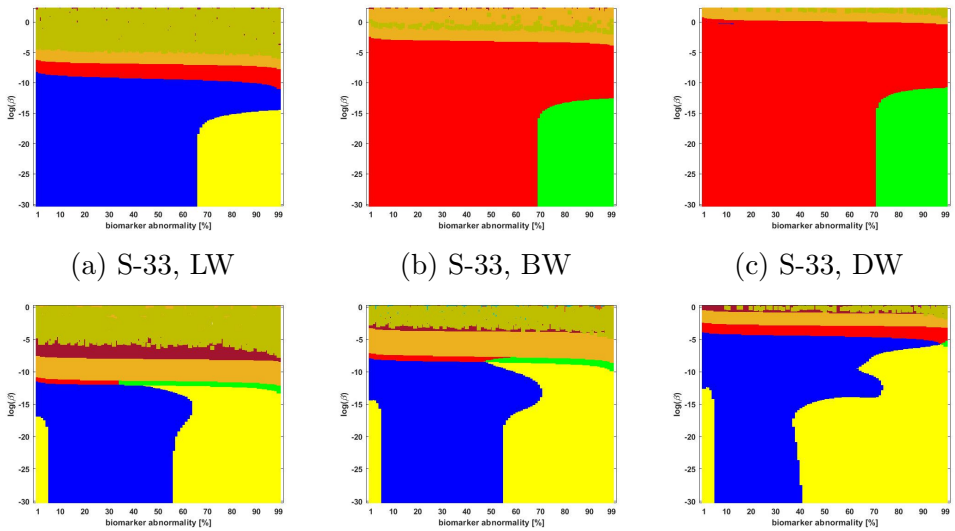


(a) S-33, LW (b) S-33, BW (c) S-33, DW



(d) S-500, LW (e) S-500, BW (f) S-500, DW

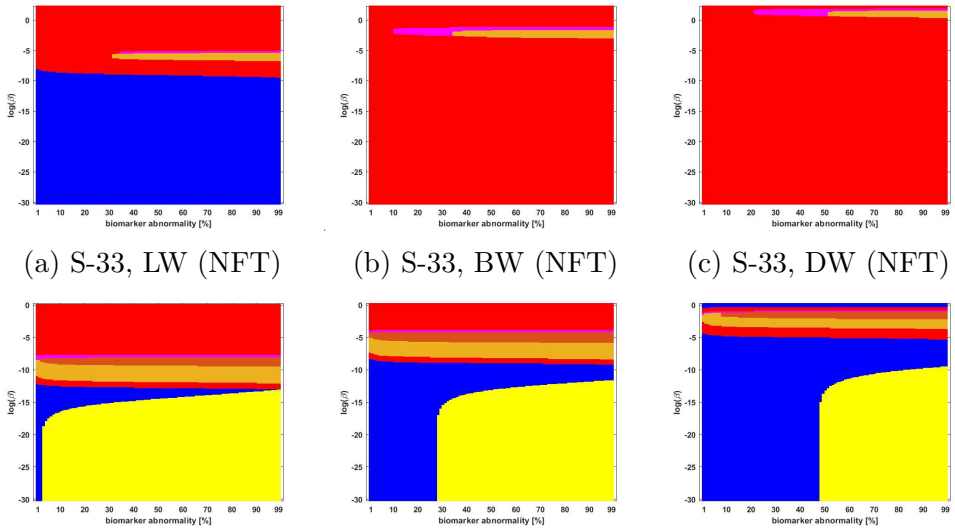
Figure 3.10: Observed computational (probabilistic) connectome τP seed staging. Noise corrected backbone at a threshold of 2.32. Axes coincide with those of Figure 3.7



(a) S-33, LW (b) S-33, BW (c) S-33, DW

(d) S-500, LW (e) S-500, BW (f) S-500, DW

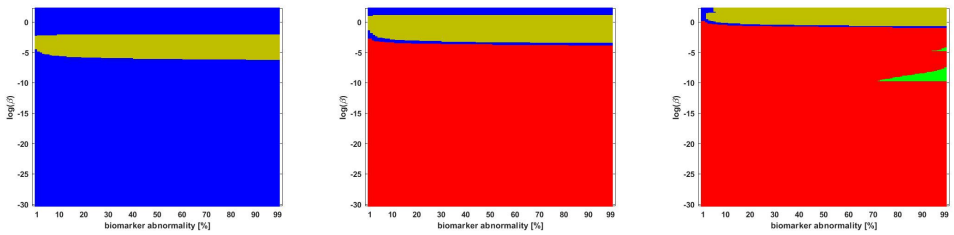
Figure 3.11: Observed computational (probabilistic) connectome τP seed staging. Naive thresholding at a threshold of 1×10^{-2}. Axes coincide with those of Figure 3.7



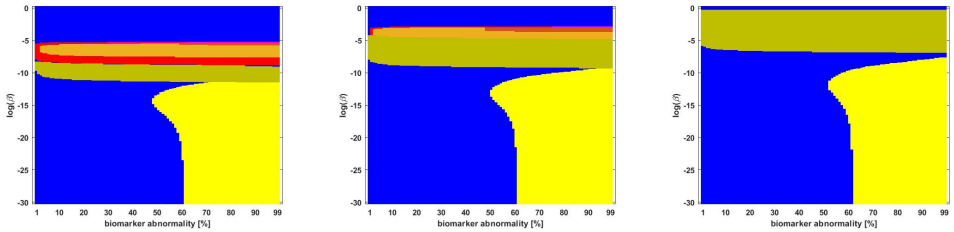
(a) S-33, LW (NFT) (b) S-33, BW (NFT) (c) S-33, DW (NFT)

(d) S-500, LW (NFT) (e) S-500, BW (NFT) (f) S-500, DW (NFT)

Figure 3.12: Observed computational (probabilistic) connectome τP NFT staging (σ=1). Density filter (DF) method at a threshold of 2×10^{-1}. The x-axis determines the biomarker abnormality threshold $1\% \leq T \leq 100\%$ and the y-axis corresponds to $-30 \leq \ln(\beta) \leq 0$ for the parameter β in (3.34b).

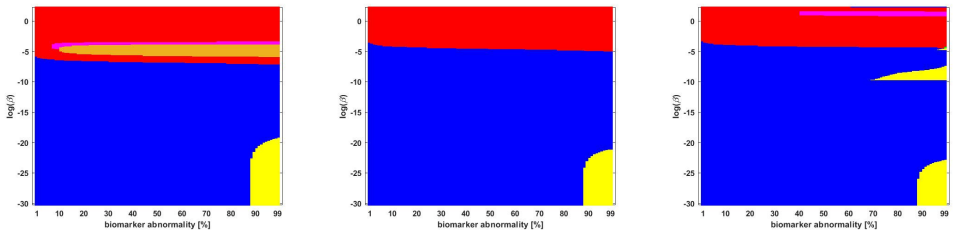


(a) S-33, LW (NFT) (b) S-33, BW (NFT) (c) S-33, DW (NFT)

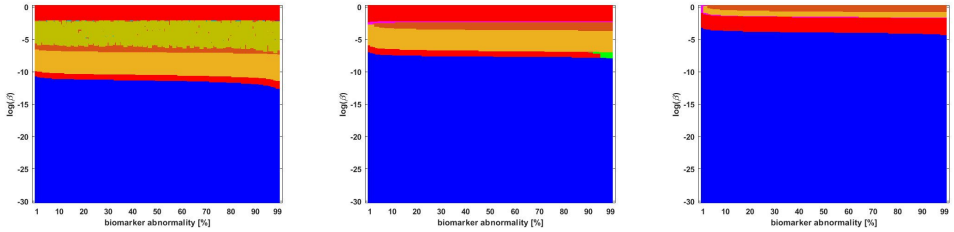


(d) S-500, LW (NFT) (e) S-500, BW (NFT) (f) S-500, DW (NFT)

Figure 3.13: Observed computational (probabilistic) connectome τP NFT staging ($\sigma = 1$). Doubly stochastic method at a threshold of 1×10^{-2}. Axes coincide with those of Figure 3.12

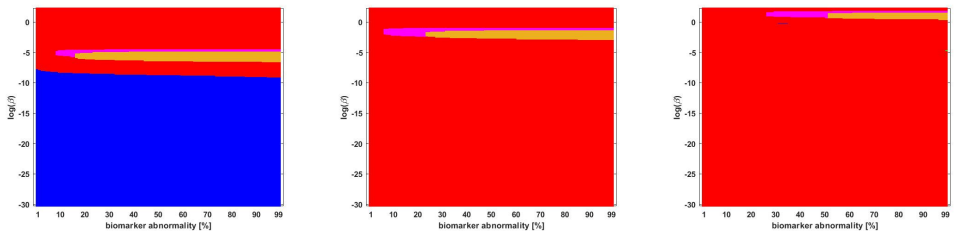


(a) S-33, LW (NFT) (b) S-33, BW (NFT) (c) S-33, DW (NFT)

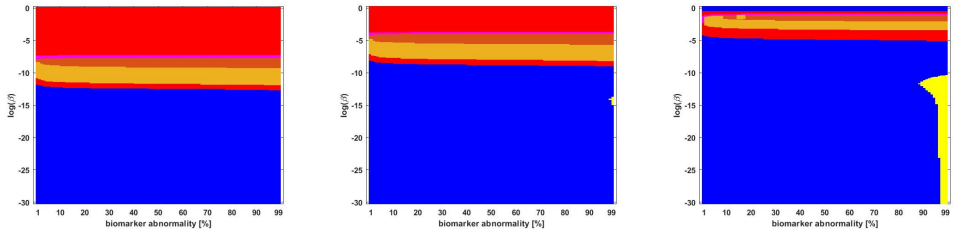


(d) S-500, LW (NFT) (e) S-500, BW (NFT) (f) S-500, DW (NFT)

Figure 3.14: Observed computational (probabilistic) connectome τP NFT staging ($\sigma = 1$). High salience skeleton at a threshold of 1×10^{-1}. Axes coincide with those of Figure 3.12

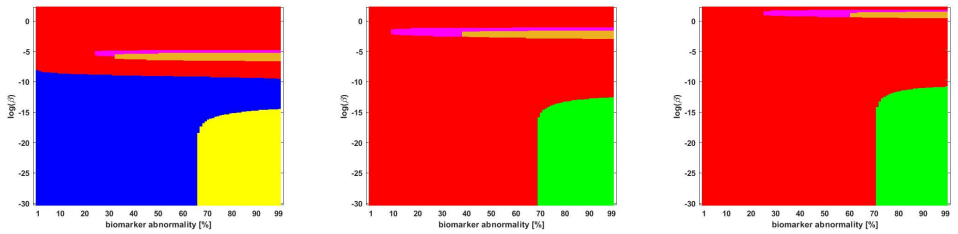


(a) S-33, LW (NFT) (b) S-33, BW (NFT) (c) S-33, DW (NFT)

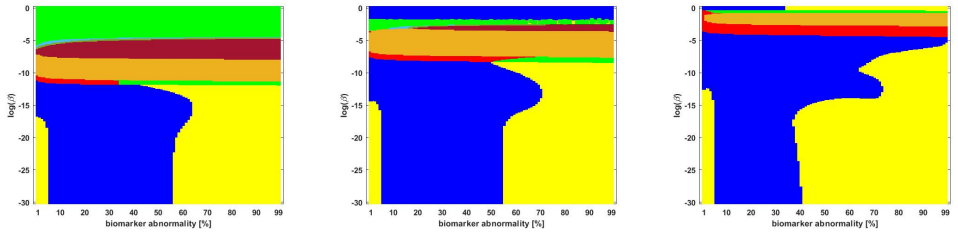


(d) S-500, LW (NFT) (e) S-500, BW (NFT) (f) S-500, DW (NFT)

Figure 3.15: Observed computational (probabilistic) connectome τP NFT staging ($\sigma = 1$). Noise corrected backbone at a threshold of 2.32. Axes coincide with those of Figure 3.12



(a) S-33, LW (NFT) (b) S-33, BW (NFT) (c) S-33, DW (NFT)



(d) S-500, LW (NFT) (e) S-500, BW (NFT) (f) S-500, DW (NFT)

Figure 3.16: Observed computational (probabilistic) connectome τP NFT staging ($\sigma = 1$). Naive thresholding at a threshold of 1×10^{-2}. Axes coincide with those of Figure 3.12

3.3.6 The role of connectome preparation in staging

In this study, we use connectomes acquired from deterministic tractography using the MRtrix software [102] and probabilistic tractography using the FSL software. We

use connectomes with multi-resolution parcellation constructed from HCP data. In general, we observe that the braid surfaces from the deterministic and probabilistic connectomes are different.

We note some similarities in the resulting staging patterns. First, both connectome types show significant sensitivity as we increase β to large values. In the diffusion-dominated regime of parameters, the observed staging patterns are look-alike with multiple overlaps for small parameter changes. Second, we find a large area of parameter space where the τP seed and NFT stagings display the same staging pattern. The similarity between τP seed and NFT predominantly occurs in the growth-dominated regime of the parameter. The last notable similarity is that both connectome types can produce the progressive Braak pattern (I $\rightarrow$ II $\rightarrow$ III $\rightarrow$ IV $\rightarrow$ V) for seed and NFT staging. We frequently observe that the progressive Braak pattern most likely emerges in the large threshold values when using probabilistic connectomes.

Despite these similarities, we also observe notable differences between the staging on the deterministic and probabilistic connectomes. The staging I $\rightarrow$ III $\rightarrow$ II $\rightarrow$ IV $\rightarrow$ V emerges clearly on the deterministic connectomes (Figures 3.5 and Figures 3.6) but is rare on the probabilistic connectomes (Figures 3.7 through 3.16). We note that the deterministic connectomes can express all four suggested staging patterns at high resolution for both τP seeds (Figure 3.5) and NFT (Figure 3.6) while the probabilistic connectomes frequently express fewer patterns.

In conclusion, connectome preparation influences a large part of model selection and can affect all downstream results.

3.3.7 The role of graph weighting scheme in observed staging paradigms

Different studies on neurodegenerative disease suggest different graph Laplacian weighting schemes. However, which weighting schemes are appropriate for modelling toxic protein transport in the brain network still needs to be investigated. Our simulations suggest that the choice of weighting schemes also affects the resulting staging patterns in three ways.

The first observed feature emerges when τP seed staging is in the diffusion-dominated regime of the parameter. The move from length-free (LW) to ballistic (BW) weights demonstrates a decreasing surface area trend sensitive to slight changes in the biomarker abnormality threshold. We also find this behaviour when moving from ballistic (BW) to diffusive (DW) weights. Figure 3.5 demonstrates the behaviour for deterministic connectomes while the top two rows of Figures 3.7 through 3.16 are

for probabilistic connectomes. We argue that the length dependence in the weights reduces the sensitivity of the τP seed staging in (3.34a) as β increases.

The second observed feature is the prevalence of the yellow staging pattern. We find that the choice of a DW graph Laplacian tends to increase the occurrence of the progressive Braak staging (I $\rightarrow$ II $\rightarrow$ III $\rightarrow$ IV $\rightarrow$ V) as shown in Figure 3.6 and the bottom two rows in Figures 3.13 through 3.16. However, we note that the emergence of the progressive Braak pattern can depend on the choice of thresholding technique. This dependence is observed in Figure 3.12 (bottom two rows) when we use the density filter method for probabilistic connectomes where the LW demonstrates this effect.

The third observed feature is the prevalence of the I $\rightarrow$ III $\rightarrow$ II $\rightarrow$ IV $\rightarrow$ V staging pattern. When we use deterministic tractography, the prevalence of this staging decreases when moving from LW to BW and again to DW. However, we note that a number of undesirable patterns emerge on the highest resolution connectome as depicted in Figure 3.6c, especially for LW when $\ln(\beta) > 0$. For probabilistic connectomes, we highlight that the choice of weights shows a weak impact on the green staging emergence and depends more on the thresholding technique and graph resolution.

To summarise, we are able to draw out the practical role of weight choice in the absence of an apparent biological reason. We will discuss more regarding the green staging pattern in the next section as it corresponds to the staging pattern observed from the SUVR data.

3.3.8 The effects of connectome resolution on observed staging

Different studies using different resolutions of connectome[145, 49, 150]. However, we wonder about the effect of connectome resolution. We study the aspect of various resolutions and observed computational staging patterns. We provide our comparative analysis according to the tractography method.

In the case of deterministic connectomes, we find that increasing the connectome resolution improves the overall observed staging results. For instance, Figure 3.5 shows a stabilising effect on τP seed staging when $\ln(\beta) > 0$ as the connectome resolution increases. Moreover, the combined area of I $\rightarrow$ II $\rightarrow$ III $\rightarrow$ IV $\rightarrow$ V and I $\rightarrow$ III $\rightarrow$ II $\rightarrow$ IV $\rightarrow$ V stagings tend to increase as the resolution increases. On the other hand, the I $\rightarrow$ II $\rightarrow$ III $\rightarrow$ V $\rightarrow$ IV and I $\rightarrow$ III $\rightarrow$ II $\rightarrow$ V $\rightarrow$ IV stagings

are more pronounced at lower resolutions. We note that the behaviour also occurs in NFT staging (Figure 3.6).

In the case of probabilistic connectomes, there is a complex relation between the thresholding method and the effect of resolution as shown in Figures 3.12 through 3.16. In general, the resolution strongly influences the resulting observed staging. However, there is no association between the thresholding method and resolution because each method has its particular effect. The low-resolution connectomes frequently exhibit a single staging pattern (typically either I $\rightarrow$ II $\rightarrow$ III $\rightarrow$ V $\rightarrow$ IV or I $\rightarrow$ III $\rightarrow$ II $\rightarrow$ V $\rightarrow$ IV) in the growth-dominated regime of the parameter. Increasing the connectome resolution may result in progressive Braak staging (I $\rightarrow$ II $\rightarrow$ III $\rightarrow$ IV $\rightarrow$ V) emergence, particularly in the large threshold values.

Overall, our results suggest that connectome resolution also determines the landscape of observable staging patterns.

3.4 Validation using data

We have discussed the model selection based on the staging problem where we use the braid surface to assess all possible observed staging patterns. A qualitative observation using simple diffusion and Fisher-KPP demonstrate the staging of τP seed and NFT on the connectome. We may obtain the progressive Braak staging pattern from different parameter values. For instance, simulations using deterministic connectomes show the Braak staging in the growth-dominated regime. This observation indicates that τP spreading can occur at different rates but show the same staging pattern. It is crucial to understand the actual τP progression from patients rather than from a theoretical point of view only. In this section, we assess potential staging sequences using flortaucipir tracer (AV-1451) data to compare with the previous results. Our task here is to analyse the progression sequence from τP PET SUVR scores according to the five Braak stages used in this work.

3.4.1 Data

Our study uses longitudinal PET data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. We use preprocessed data containing 1,184 records where each record associates with a subject structural MRI and PET scans. The dataset are originally obtained from 786 patients, with 403 female and 383 male patients. The age average of overall subjects is 71.32 years with a 6.75 years deviation, with a female age average of 70.26 years with a 6.68 years deviation and a male age average of

72.45 with a 6.65 years deviation. ADNI provide subject classification in their τP PET scans where patients are categorised into cognitively normal (CN), significant memory concern (SMC), mild cognitive impairment (MCI) with additional early (EMCI) and late (LMCI) MCI, and AD groups. The CN and SMC groups are actual patients with normal cognitive function. However, patients in the SMC declared that they believe having a memory issue. Of these, there are 572 CN records from 397 patients, 92 SMC from 49 patients, 138 EMCI from 84 patients, 222 MCI from 164 patients, 75 LMCI from 38 patients, and 86 AD records from 54 patients. The data are available on the ADNI website (adni.loni.usc.edu) and are preprocessed at the Helen Wills Neuroscience Institute at the University of California, Berkeley. The compiled results are freely available through the same website. They provide a companion document elaborating on the preprocessing steps along with the data.

The Berkeley preprocessing pipeline is briefly explained as follows. Each patient T1 MRI was segmented using FreeSufer software 7.1.1 and each normalised-intensity flortaucipir scan was co-registered to their bias-corrected T1 image for each patient. Then, the partial volume effects were corrected using the geometric transfer matrix approach [9, 10]. The dataset contains flortaucipir uptake for each region according to the FreeSurfer segmentation. Before analysing the data, we normalised each subject's regional SUVR score using the inferior cerebellar grey matter as the reference region, as suggested in their documentation.

We create two datasets using this preprocessed dataset for our data analysis. We compute the whole brain volume-weighted SUVR average. This dataset is a global score for each patient record. Next we compute the regional volume-weighted SUVR average for each patient. As a result, we have the *base dataset* where each patient has a record with a global SUVR score and five volume-averaged regional scores associated with the regions of Table 3.2. This dataset is the first dataset. The second dataset comes from clustering each patient record into five groups of Braak stage in the next section.

3.4.2 Clustering methods

We use Conditional Inference Tree (CIT) algorithm to cluster the patient dataset into five groups of the Braak stage. The CIT algorithm is a kind of decision tree algorithm that uses recursive partitioning steps of dependent variables based on their correlation values [58, 77]. It is similar to the decision trees but uses a significance test. The test is a permutation test that selects a covariate to split and recurse the variable. The method calculates the p-value in this test. The test is for selecting input variables

Stage group classification	Volume-averaged global SUVR (GS)	Subject records in group
I	$1.266 \leq$ GS	36
II	$1.266 <$ GS ≤ 1.392	248
III	$1.392 <$ GS ≤ 1.52	443
IV	$1.52 <$ GS ≤ 2.365	416
V	GS > 2.365	41

Table 3.4: Conditional inference-based partitioning of ADNI flortaucipir data into the stages of Table 3.2

rather than maximising the information measure. The algorithm is briefly described as follows: (1) It performs a test of the global null hypothesis between random input and response variables, (2) it then selects the input variables with the optimal p-value of response variables, (3) it then performs a binary split on the selected input variables, and lastly (4) it recursively performs step (1) - (3) until the number of desired clusters is constructed. This algorithm has a drawback: it is unsuitable for data with missing values.

We now create a second dataset called *partitioned dataset* by following the methodology of [151]. We follow the identical approach in [151]for our application of CITs. The objective is to classify each patient record into one of the pre-defined regional stages. First, we use the global score for a decision variable. Second, the CIT algorithm produces a global score threshold, separating the Braak V regional volume-weighted SUVR from the other stages. Third, the process is repeated for the remaining global score producing a new threshold and separating the Braak IV regional volume-weighted SUVR score from the other three stages. We repeat the algorithm until we obtain five different clusters of the Braak stage. Thus, we have five separate files corresponding to Table 3.2. All partitions are significantly different (p-value<0.05) under a Welch's t-test. Table 3.4 shows the partitioned dataset and the thresholds produced by the CIT algorithm.

3.4.3 Data analysis

In this part, our goal is to investigate whether the ADNI data suggest alternative staging patterns for the computational model selection and staging problem. We use both datasets to perform several statistical analyses. We compute correlation values between regions, regression coefficients with respect to gender and age, and Z-score to evaluate the staging sequence from τP PET data.

Patient group and region	Connectome Braak regional SUVR correlation				
	Region I	Region II	Region III	Region IV	Region V
Group I, Region I	1.0*	0.26	0.53*	0.14	0.04
Group II, Region II	0.34*	1.0*	0.31*	0.14*	-0.29*
Group III, Region III	0.69*	0.32*	1.0*	0.06	-0.09
Group IV, Region IV	0.11*	0.28*	0.32*	1.0*	-0.13*
Group V, Region V	-0.13	-0.03	0.16	-0.08	1.0*

Table 3.5: Pearson correlation between a stage group's primary region and all other regions within the group. A $*$ denotes a statistically significant correlation ($p < 0.05$) between regions.

First, we compute the Pearson correlation between each stage's primary region and all other regions within that stage group. Table 3.5 depicts the results. Assuming that NFT pathology begins in Region I, the correlation value suggests that staging sequences with I $\rightarrow$ III may also be of potential for NFT. Assuming that Region III continues Region I, III $\rightarrow$ II could be the next suggested progression. This investigation suggests that sequences starting with the prefix I $\rightarrow$ III $\rightarrow$ II could be another potential other than I $\rightarrow$ II $\rightarrow$ III. Finally, there is no significant correlation value in Group V, likely due to sparse data in that group.

Second, we use regression analysis to compare each regional volume-averaged SUVR value to that of the other regions. Table 3.6 shows the results. We evaluate evidence for influence using standardised factor scores for significant factors. The highest R^2 and B show that Region I strongly affects Region III ($R^2 = 0.775$, $B = 0.613$, $p < 0.001$). This finding supports our previous observation using correlation values. Region I also influences Region II ($R^2 = 0.486$, $B = 0.486$, $p < 0.001$) with a lower value of both B and R^2 than Region III. However, Region III affects Region II more than Region I as shown by the B score ($R^2 = 0.42$, $B = 0.5$). We observe that the variational fidelity (R^2) of the Region III model decreases by 24.8% and the Region II model fidelity by 13.6% when removing Region I. This investigation suggests that a I $\rightarrow$ III $\rightarrow$ II prefix is also a possible outcome of computational staging when assuming Region I as the origin of pathology. We investigate the staging suffix options between IV $\rightarrow$ V and V $\rightarrow$ IV. We use the first three regional scores in the regression for Region IV and Region V. Region III affects most Region IV ($R^2 = 0.43$, $B = 0.75$, $p < 0.001$) compared to Region I and Region II. Region V shows a lower score than Region IV indicating that Region III marginally describes Region V ($R^2 = 0.22$, $B = 0.38$, $p < 0.001$). The investigation using the data suggests

Regression model	Influential model region(s)	Model R^2	B	p-value
Region II	Region I	0.486	0.49	$p < 0.001$
	Region III	0.42	0.5	$p < 0.001$
Region III	Region I	0.78	0.61	$p < 0.001$
	Region IV	0.58	0.39	$p < 0.001$
	Region II		0.36	$p < 0.001$
Region IV	Region III	0.43	0.75	$p < 0.001$
Region V	Region III	0.22	0.38	$p < 0.001$

Table 3.6: Regression models, covarying for age and gender, of regional influence

Region	% Involved	Z
Region I	100.0	3.24
Region III	55.1	2.55
Region II	46.94	1.81
Region IV	22.45	1.43
Region V	22.45	1.17

Table 3.7: Spreading order of τP (SUVR) for Region-I involved patient records

that a IV $\rightarrow$ V may be more acceptable than a V $\rightarrow$ IV suffix. This result supports the conclusion from Table 3.5.

Third, we now verify our previous findings. We use a similar approach to [31] and compute Z-scores from the base dataset for the regions in Table 3.2. Similar to [31], we standardise all regional mean SUVR scores and consider a particular region in a patient record if that region's Z-score is $Z \geq 2$. We select all patient records with an involved Region I (entorhinal cortex) and continue with this subset data for computing all other involved regions' mean Z-score. Table 3.7 shows the suggested ordering of regions by the SUVR data. The last finding confirms that the progression sequence I $\rightarrow$ III $\rightarrow$ II $\rightarrow$ IV $\rightarrow$ V may be a suggested pattern.

We summarise our validation section. We have computed the progression sequence from the ADNI SUVR data using correlation values, regression analysis, and Z-score ordering. Our findings suggest that I $\rightarrow$ III $\rightarrow$ II $\rightarrow$ IV $\rightarrow$ V is a pattern of interest for τP NFT and possibly for τP seed. This sequence is represented by the green staging colour in the Braid surface. For computational staging, we conclude the suggested patterns in Table 3.8. However, this result can be still extended by adding more late-stage data, especially Group V, to confirm a higher degree of certainty in the whole computational τP staging for AD.

Computational staging patterns (see Table 3.2)		
Canonical late staging	I $\rightarrow$ II $\rightarrow$ III $\rightarrow$ IV $\rightarrow$ V*	I $\rightarrow$ III $\rightarrow$ II $\rightarrow$ IV $\rightarrow$ V†
Uncertain late staging	I $\rightarrow$ II $\rightarrow$ III $\rightarrow$ V $\rightarrow$ IV‡	I $\rightarrow$ III $\rightarrow$ II $\rightarrow$ V $\rightarrow$ IV‡

Table 3.8: Computational staging patterns of potential interest for τP seeds and NFT. * Progressive Braak staging, † Suggested by SUVR data, ‡ Additional connectome computational stagings of potential interest pertaining to uncertain τP seed staging prefixes.

3.5 Concluding remark

In this chapter, we have presented methods to address the staging problem that occurs in modelling toxic protein propagation throughout the brain network. We aim to capture the qualitative staging of τP protein as observed in Alzheimer's disease. We developed tools for model selection of toxic protein transport problem and compared the resulting patterns from the model against patterns from data.

Studies of neurodegeneration conducted on the diffusion process have uncovered notable features of brain atrophies [145]. However, the diffusion considered in these studies does not correspond to a physical transport process and satisfies neither mass conservation nor Fick's condition. We have shown that physical diffusion can only be satisfied by the standard Laplacian form $\mathbf{L} = \mathbf{D} - \mathbf{A}$.

We introduced in this chapter braid diagrams and braid surfaces for analysing staging. These tools present a direct and visual assessment of a complex nonlinear process evolving in time. The braid diagram allows one to assess staging when a particular parameter is used. Meanwhile, the braid surface demonstrates a large number of possible patterns from the selected model.

Simulation results have demonstrated that τP seeding and its accumulation propagate systematically in structural connectomes identified by the desirable Braak staging similar to empirical observations. However, the desirable pattern obtained from our models demonstrates that parameter selection, connectome weight, and the brain's underlying graph strongly influence the Braak staging occurrence in model simulations. In addition, the connectome is a data-driven model affected by the acquired network streamlines.

The generalised staging problem in AD poses novel theoretical questions and shows the need for more experimental data. A further study of τP seed staging patterns could help to reveal a more plausible selection of parameter values. We believe that more PET data for patients in the higher Braak stages would help to reduce uncertainty in exploring AD progression patterns. Moreover, it is important to investigate

the relationship between the connectome topology, the model parameters, and the weighted distances between ROIs for theoretical research purposes.

Chapter 4

Arrival-Time Problem and Applications to Neurodegenerative Disease

4.1 Overview

4.1.1 Previous work on Braak staging

Braak staging is a system for identifying the progression of Alzheimer's disease through toxic τP spreading in the brain, with each stage being associated with a particular set of brain regions. This propagation feature can be reproduced through mathematical models that show computational regional spreading. The previous chapter has discussed possible resulting patterns from a mathematical network model and validation using data from τP PET scans. The results suggest that there are alternative staging patterns and different regimes of propagation. Despite the important findings of the previous chapter, we still inquire about the underlying systematical propagation on the brain network.

The simplest model that we use to study such structured staging sequence on an undirected connected network is a discrete version of the Fisher-Kolmogorov-Petrovsky-Piskunov (Fisher-KPP) reaction-diffusion equation [188]. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a connected and undirected network with a collection of N nodes $\mathcal{V}$, representing regions of interest, and edges $\mathcal{E}$ representing connections between these regions. If $c_i(t)$ is the quantity of interest in a region i, modelled as a node in a network, and evolving over time t, then on this network, the dynamics is governed by a system of

N ordinary differential equations with N initial conditions

$$\begin{aligned} \frac{\mathrm{d}c_i}{\mathrm{d}t} &= -\rho \sum_{j=1}^{N} L_{ij} c_j + \alpha c_i (1 - c_i), \quad i = 1, \ldots, N, \\ c_i(0) &= q_i, \quad i = 1, \ldots, N, \end{aligned} \tag{4.1}$$

where parameters ρ and α are positive and denote the diffusion constant and growth constant, and $q_i \in [0, 1]$ for all i. The nonlinear kinetics is coupled with a linear diffusion on a network represented by the symmetric *graph Laplacian* [119] $\mathbf{L}$

$$L_{ij} = -A_{ij} + \delta_{ij} \sum_{k=1}^{N} A_{ik}, \tag{4.2}$$

where $\delta_{ij} = 1$ if $i = j$, and 0 otherwise. The graph Laplacian encodes the connection between regions $L_{ij} = 0$ if there is no connection between different nodes i and j.

In this chapter, we address questions to understand the overall dynamics of an autocatalytic process propagating on a network through *the arrival-time problem*. 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? Motivated by the study of toxic protein propagation in neurodegenerative diseases, we present and compare three different estimates for the arrival time. We use the Fisher-Kolmogorov-Petrovsky-Piskunov equation as a paradigm for the dynamics and show that each method provides different insight and time estimates. Specifically, 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. Our study also highlights the accuracy of the three methods of producing staging patterns.

4.1.2 Regarding arrival times and methods

The notion of arrival times has become an essential problem, especially in the field of epidemiology to model disease spreading on a large-scale network connecting various regions. This framework allows one to study the early spread of a disease through a sequence of times instead of common frequencies such as the number of infections and density. Doing so, we can rank the spreading between regions according to the arrival of the disease and systematically investigate the epidemic spread on a large heterogeneous network [57].

There are two main approaches to obtaining arrival times. One can start with a stochastic random walk on a graph and study its mean-field approximation [23, 57, 80].

The other approach is to see the model as a continuum dynamical process that is independent of any discrete process [152, 197]. Despite the difference in perspectives, we note that there are three useful methods to derive the arrival times.

First, one can consider the equivalent travelling fronts propagating on a simple topology network such as an infinite homogeneous tree or a periodic tree [75, 199]. The natural approach is to linearise the system around the unstable zero state under the assumption of *linearly determinate system* where the full nonlinear system spreading speed equals to its linearised system. This approach allows one to obtain the linear velocity of propagation. However, such an approach is applicable to limited systems, particularly to simple and homogeneous networks, although the application of this approach gives many exploratory results on lattices [48] and elucidates some counterintuitive phenomena in disease spreading [14].

Second, one can linearise the system around an equilibrium state. Then, the variable solution at each node can be obtained as the solution of this linear system for a given initial condition. From the solution, the arrival times are given as the solution of a transcendental equation at each node [30, 89]. The advantage of this approach is that the linearisation provides an adequate estimate for small initial concentrations and large diffusion. Yet, its main disadvantage is the method does not provide much insight into the dynamics.

The third approach is to consider the propagation between two nodes along a determined path. Again, we initially linearise the system to extract the time it takes for the second node to reach the same level as the initial condition in the seed node [57] while neglecting all other connections. There are two perspectives that we can use to obtain arrival times. First, one can consider this propagation along the path in terms of the shortest path and solve the linearised system along that path to obtain arrival times. Second, one can extract independent symmetric arrival times between two nodes neglecting other connections attached to those nodes. It allows one to obtain arrival times between any two nodes on a network through a sum along a path, where the shortest path provides the notion of distance called *effective distance* [23]. Moreover, one can also consider paths other than the shortest path in the existence of multiple paths on that network at the cost of extra computation [80].

4.1.3 Arrival time problem

We start with the *canonical arrival-time problem*, in which 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 $\tau_i = \tau(\mu)$ such that $c_i(\tau_i) = \mu$ (with $\beta \leq \mu < 1$ so

that $\tau_i \geq 0$ and finite with the equality $\tau_i = 0$ attained if only if $i = s$ and $\mu = \beta$). The *generalised 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 $\tau_i = \tau(\mu)$ such that $c_i(\tau_i) = \mu$ (with $\beta \leq \mu < 1$ so that $\tau_i \geq 0$ for all i).

4.2 Methods

In epidemiology, arrival times are the times when a disease emerges in regions that are not the source of the disease. Previous studies explored the notion of arrival times mostly through a linear approach. The estimates derived in this study are based on the Fisher-KPP dynamics 4.1. In this case, we linearise the system around the healthy state and it reduces to,

$$\begin{aligned} \frac{\mathrm{d}c_i}{\mathrm{d}t} &= -\rho \sum_{j=1}^{N} L_{ij} c_j + \alpha c_i, \\ c_i(0) &= q_i, \quad i = 1, \ldots, N. \end{aligned} \tag{4.3}$$

Let $\mathbf{c}(t) = (c_1, ..., c_N)^T$, $\mathbf{q} = \mathbf{c}(0) = (q_1, ..., q_N)^T$, and $\mathbf{M} = \alpha\mathbf{I} - \rho\mathbf{L}$, where $\mathbf{I}$ denotes the identity matrix. Then, the concentration of every node at time t is given by

$$\mathbf{c}(t) = e^{t\mathbf{M}}\mathbf{q}. \tag{4.4}$$

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, $i \in \{1, \ldots, N\}$ so that

$$\mathbf{L}\mathbf{v}^{(i)} = \lambda_i \mathbf{v}^{(i)}, \quad i = 1, \ldots, N. \tag{4.5}$$

Further, by construction, we have $\lambda_1 = 0$ with $\mathbf{v}^{(1)} = (1, \ldots, 1)$ and, since the graph is connected, $0 < \lambda_i \leq \lambda_j$ for all $1 < i < j \leq N$ [33]. Then, the linear solution can be written explicitly as

$$\mathbf{c} = e^{\alpha t} \sum_{i=1}^{N} \left(\mathbf{q} \cdot \mathbf{v}^{(i)} e^{-\rho\lambda_i t}\right) \mathbf{v}^{(i)}. \tag{4.6}$$

Figure 4.1 illustrates the Fisher-KPP evolving on a 5-node network in two regimes of interest as shown. Let λ_2 be the smallest nonzero eigenvalue of $\mathbf{L}$. The *diffusion-dominated regime* is defined by $\rho\lambda_2/\alpha \gg 1$ and the *growth-dominated regime* is defined by $\rho\lambda_2/\alpha \ll 1$. We first explore the diffusion-dominated case.

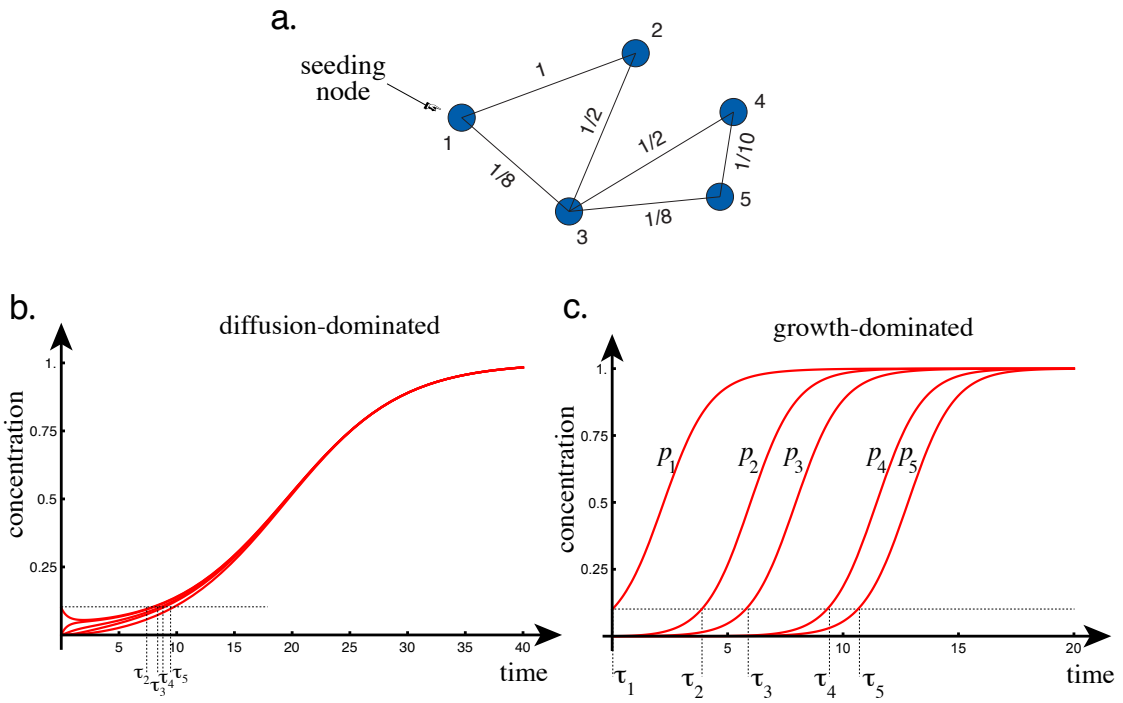


Figure 4.1: Example of dynamics on a 5-node network (a). Initially, only the first node is seeded ($c_1(0) = 1/10$). In the diffusion-dominated case (b: $\alpha = 1/5, \rho = 1$), all concentrations quickly approach $1/10$ before experiencing the effect of the growing exponential in unison. For the growth-dominated case, after growth at the first node, the second node is invaded and then all other modes are subsequently invaded in a front-like progression (c: $\alpha = 1, \rho = 1/100$).

We note that if diffusion dominates, then the early dynamics is governed by the slowest dynamics in (4.6), 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

$$\frac{\mathrm{d}c}{\mathrm{d}t} = \alpha c(1-c), \quad c(0) = \mu/N. \tag{4.7}$$

The solution of this equation is,

$$c(t) = \frac{\mu e^{\alpha t}}{N - \mu + \mu e^{\alpha t}}, \tag{4.8}$$

with an arrival time at all nodes, when $c(\tau_{\text{diff}}) = \mu$, given by

$$\tau_{\text{diff}} = \frac{1}{\alpha} \log\left(\frac{N-\mu}{1-\mu}\right). \tag{4.9}$$

We can also obtain the approximation of diffusion-dominated arrival time by assuming $\rho\lambda_i/\alpha \to \infty$ for $i \neq 1$. The concentration is predominantly determined by the graph Laplacian's Kernel $\mathbf{v}^{(1)} = (1, \ldots, 1)$ while other modes vanish since $\alpha - \rho\lambda_i < 0$ for $i \neq 1$. Hence, we have $\mathbf{c}(t) = \mu/Ne^{\alpha t}\mathbf{v}^{(1)}$ where $q_s(0) = \mu$ at the seed node s and otherwise zero. Therefore, the arrival time of all nodes is given by

$$\hat{\tau}_{\text{diff}} = \frac{1}{\alpha} \ln(N). \tag{4.10}$$

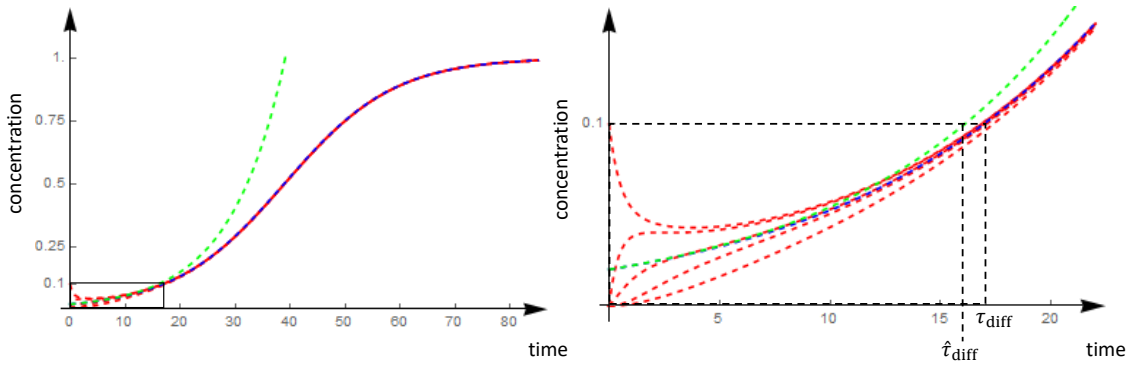


Figure 4.2: Diffusion-dominated case for the 5-node network with $\alpha = 1/10, \rho = 1, c_1(0) = \mu = 1/10$. We see that the dynamics of a single node with initial condition $c(0) = \mu/N$ (blue dashed line) provides an excellent approximation of the entire solution (left) and, looking at the close-up in the black rectangle, for the arrival time as well (right).

Figure 4.2 shows the comparison between the exact solution (red dashed lines), diffusion-time solution (blue dashed line), and its linearisation (green dashed line). We note that the diffusion-time solution becomes a good estimate at the threshold $\mu = 1/10$. As an additional note, one can obtain the approximation when $\rho\lambda_2/\alpha \approx 1$ as:

$$c_i(t) = \mu/Ne^{\alpha t} + \mu\left(v_i^{(2)}\right)^2. \tag{4.11}$$

However, this estimate does not give a better estimate compared to the diffusion-time solution or even its approximation (not shown).

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

4.2.1 Linear estimate

The first approach to the arrival time in the growth-dominated regime is the linear time. The *general linear arrival time* $\widehat{\tau}_j$ is given by the solution of the equation

$$\mu = e^{\alpha\hat{\tau}_j} \sum_{i=1}^{N} \left(q \cdot q^{(i)} e^{-\rho\lambda_i \hat{\tau}_j}\right) q_j^{(i)}. \tag{4.12}$$

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

$$\mu/\beta = e^{\alpha\hat{\tau}_j} \sum_{i=1}^{N} \left(q_j^{(i)} q_s^{(i)} e^{-\rho\lambda_i \hat{\tau}_j}\right). \tag{4.13}$$

Figure 4.3 demonstrates the exact solution (red dashed lines) and the linear approximation (green dashed lines). We note that the linear solution increases faster than the exact solution, which gives smaller estimates as a consequence.

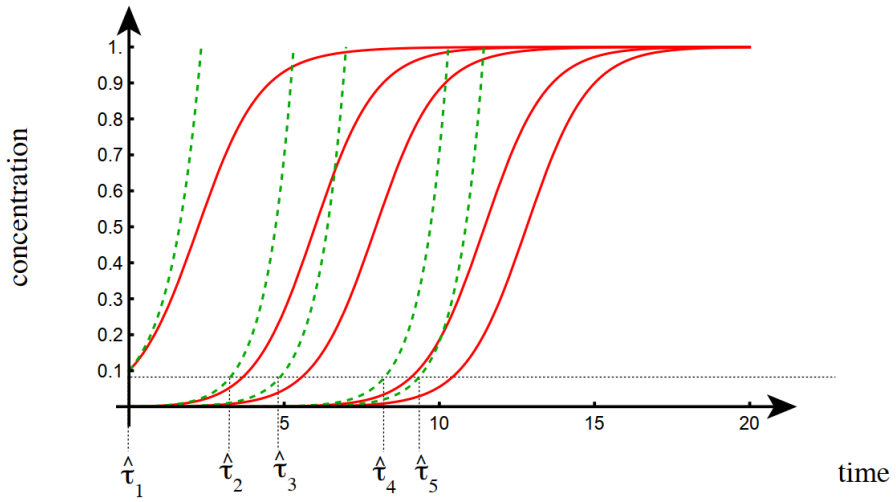


Figure 4.3: a: Comparison between the exact (red) and linear solutions (dashed, green) for the 5-node network ($\alpha = 1, \rho = 1/100, \beta = 1/10$).

To illustrate the performance of the linear times we compute the error made as a function of the seed size:

$$\mathcal{E}_{\text{lin}}(\mu) = \sum_{i=1}^{N} |\hat{\tau}_i(\mu) - \tau_i(\mu)|. \tag{4.14}$$

Fig. 4.4 gives a comparison of the linear solutions with the numerical solutions for our 5-node network. We observe the following features. First, the approximation given by the linear arrival times improves as μ decreases. This behavior is illustrated

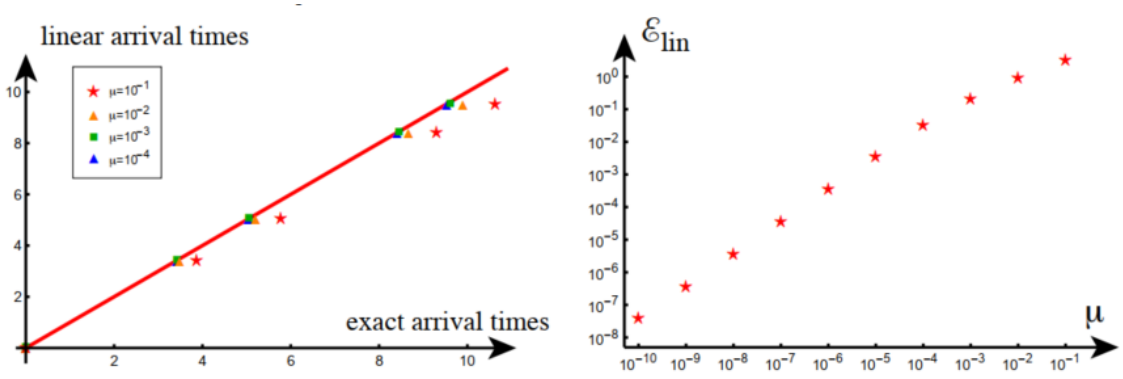


Figure 4.4: a: Comparison between the exact arrival times vs linear arrival times for various initial seeding (left) and error between exact and linear arrival times (right) as a function of initial seeding shows good convergence, as expected.

in Fig. 4.4 where we compute the error. As $\mu \to 0$ with $\mu > 0$, this approximation converges to the exact value as the nonlinear terms become negligible.

Second, since the nonlinear terms of (4.1) act to reduce the value of the concentration in time, and (4.3) 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 obtain actual values of the linear arrival times we need to solve N individual transcendental equation of the form (4.6), one for each arrival time. Solving this system can prove to be numerically challenging without a decent first guess.

Fourth, this method is often overlooked in favour of more involved graph-based measures based on the superposition of multiple paths. However, the exponential matrix solution (4.4) naturally combines the sum of all paths as expected from a diffusion operator.

Overall, the linear arrival times constitute a robust universal method valid for small seed concentrations to obtain arrival times.

4.2.2 Lambert estimate

4.2.2.1 Two-node model

A second approach is to define a 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

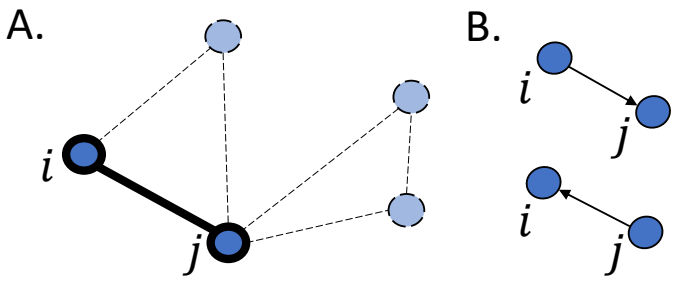


Figure 4.5: a: Two-node model that is neglecting the other connections. b: Independent two-node model with an arbitrary direction of propagation.

approximation (4.3) with a surrogate graph Laplacian, corresponding to the two-node subnetwork of node i and j, given by

$$\mathbf{L}_{ij} = \begin{bmatrix} A_{ij} & -A_{ij} \\ -A_{ij} & A_{ij} \end{bmatrix}. \tag{4.15}$$

Using the eigenvectors of $\mathbf{L}_{ij}$ and (4.13) 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

$$e^{\alpha t}\left(1 - e^{-2\rho A_{ij} t}\right) = 1. \tag{4.16}$$

Expanding for small ρ, we obtain

$$\rho A_{ij} t e^{\alpha t} = 1. \tag{4.17}$$

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

$$t_{ij} = \frac{1}{\alpha} W_0\left(\frac{\alpha}{\rho A_{ij}}\right), \qquad i \neq j. \tag{4.18}$$

Figure 4.6 depicts the Lambert W_0 function when we choose $\alpha = \rho = A_{ij} = 1$ for (4.18). We note that the Lambert W_0 function is the principal branch of the Lambert function and will give a single nonnegative real root because $\alpha/\rho A_{ij} \geq 0$. We also 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} between two neighbouring nodes as the *Lambert edge distance*.

4.2.2.2 Lambert distance

From this pairwise distance between neighbouring nodes, we define the *Lambert distance* W_{ij} as the shortest path with respect to the Lambert edge distance between

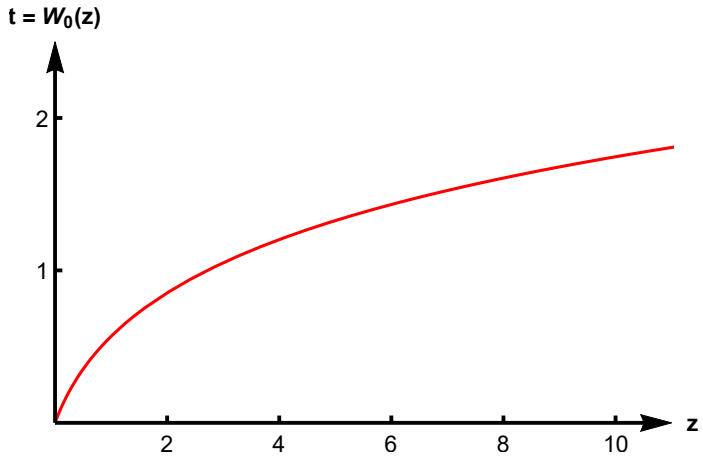


Figure 4.6: Lambert W_0 function plot of (4.18).

two nodes i and j. Let $\Gamma_{ij} = (\gamma_0, \gamma_1, \ldots, \gamma_n)$ with $\gamma_0 = i$ and $\gamma_n = j$ be this shortest path, then

$$W_{ij} = \begin{cases} \displaystyle\sum_{k=0}^{n-1} t_{\gamma_k,\gamma_{k+1}} = \frac{1}{\alpha} \sum_{k=0}^{n-1} W_0\left(\frac{\alpha}{\rho A_{\gamma_k,\gamma_{k+1}}}\right), & i \neq j, \\ 0, & i = j, \end{cases} \tag{4.19}$$

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

$$\frac{\mathrm{d}c}{\mathrm{d}t} = \alpha c(1-c), \quad c(0) = \beta, \tag{4.20}$$

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

$$t_{ii} = \frac{1}{\alpha} \log\left(\frac{(1-\beta)\mu}{(1-\mu)\beta}\right). \tag{4.21}$$

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

$$\tilde{\tau}_{ij} = t_{ii} + W_{ij}. \tag{4.22}$$

Figure 4.7 demonstrates the Lambert arrival times compared to the linear times. We see that the Lambert solutions, and Lambert arrival times, are to be an improvement over both the linear solutions, and arrival times, and can capture the dynamics

over the entire domain. The Lambert arrival times (4.19) 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.

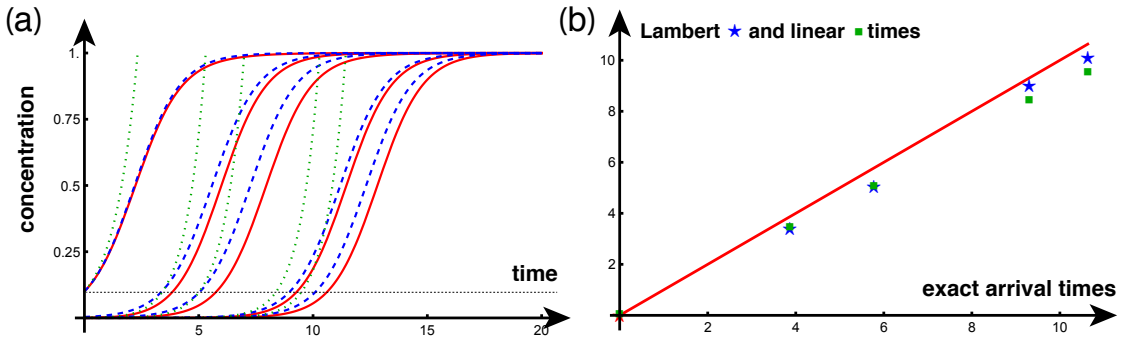


Figure 4.7: a. Exact (red), Lambert (blue, dashed), and linear (green, dotted) solutions with $\alpha = 1, \rho = 1/100, c_1(0) = \mu = \beta = 1/10$. b. Lambert arrival times (blue stars) compared to linear arrival times (green squares).

We note some important features regarding the Lambert distance. First, the method provides a good estimate without depending on the initial concentrations. We could see that the Lambert distance is an intrinsic property of the system. Second, the Lambert distance includes the role of diffusion, which has a weak influence. 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. Despite its remarkable feature and simplicity, the method is still based on linearisation. Hence, the approximation lacks of saturating effects that increase the arrival times. Overall, the Lambert distance provides a useful estimate for the arrival times.

4.2.3 Further computation on Lambert estimate

We have developed the Lambert estimate based on a single shortest path from a single seeding node. Unlike the Lambert distance, the linear estimate includes multiple paths that connect the seed and target nodes. It also allows multiple seeds in the equations and we only solve a transcendental equation for each node to compute the arrival times. In this section, we introduce the computation for multiple paths and multiple seeds for the Lambert distance.

4.2.3.1 Propagation on a line graph

We start with the notion of propagation on a line graph before we define the computation of multiple paths and multiple seeds. The process is assumed to occur similarly to the two-node case and the arrival time of the next node is close to the current node.

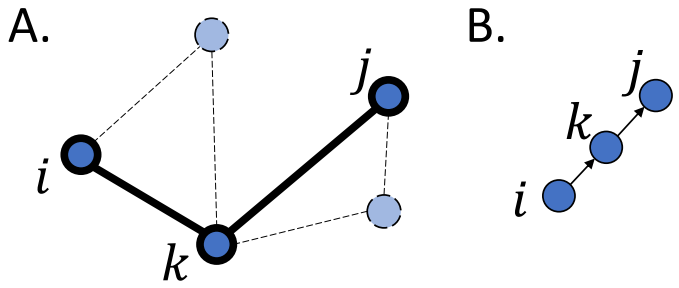


Figure 4.8: a: A line model that is neglecting the other connections. b: Independent line model with an arbitrary direction propagation.

Let $\Gamma_{ij} = \gamma_0, ..., \gamma_n$ be the shortest path that connects i and j. We can model the propagation on a line graph, as illustrated in Figure 4.8, as a system of ODEs given by,

$$\frac{\mathrm{d}c_{\gamma_1}}{\mathrm{d}t} = \alpha c_{\gamma_0}, \tag{4.23a}$$

$$\frac{\mathrm{d}c_{\gamma_{k+1}}}{\mathrm{d}t} = \alpha c_{\gamma_{k+1}} - \rho L_{\gamma_k,\gamma_{k+1}} c_{\gamma_k}, \qquad k = 1, ..., n-1, \tag{4.23b}$$

with $\gamma_0 = i$ and $\gamma_n = j$. The solution[4] at node j is given by

$$c_j(t) = \mathcal{T}(\Gamma_{ij}; n) \frac{(\rho t)^n}{n!} \beta e^{\alpha t}, \tag{4.24}$$

where $\mathcal{T}(\Gamma_{ij}; n) = \prod_{k=0}^{n-1} A_{\gamma_{k+1},\gamma_k}$ denotes the multiplication of adjacency weights on the shortest path of length n. When $\mu = \beta$, the arrival time is given by

$$t_{ij} = \frac{n}{\alpha} W_0 \left(\frac{\alpha}{n\rho} \sqrt[n]{\frac{n!}{\mathcal{T}(\Gamma_{ij}; n)}} \right). \tag{4.25}$$

Figure 4.9 demonstrates the influence of parameter to the resulting t_{ij}. The arrival time t_{ij} increases as n goes larger indicating that a node with further shortest path

[4]One can find the solution by defining a surrogate Laplacian of the shortest path first. Following the same assumption as the two-node model, one can use the eigenvector of the surrogate graph Laplacian $\mathbf{L}_{ij}$. Expanding the solution for small ρ and using $\mu = \beta$, one will obtain the same expression of concentration $c_j(t)$ (4.24) and obtain the arrival time (4.25).

length from the seed will have a longer arrival time. If we fix two parameters, the arrival time decreases as the free parameter increases (Figure 4.9a-c). The influence of two free parameters is depicted in Figure 4.9d-f. From the contour plots, we note that the arrival time of nodes $n = 1$ reduces faster than nodes with $n = 2$. This observation is shown by the larger gap between two threshold curves for the same n.

For the rest of the section, we will use this perspective as our foundation to obtain arrival times for multiple seeds and multiple paths.

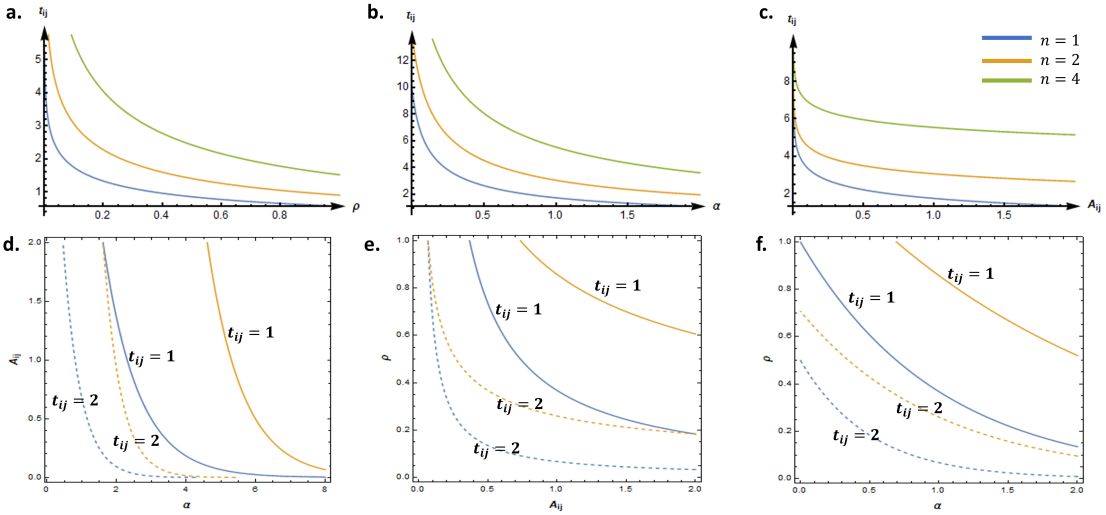


Figure 4.9: t_{ij} plots for $n = 1$ (blue), $n = 2$ (orange), $n = 4$ (green) as a function of: (a) ρ when $\alpha = A_{ij} = 1$; (b) α when $A_{ij} = 1$ and $\rho = 0.1$; (c) A_{ij} when $\alpha = 1$ and $\rho = 0.1$. Sensitivity analysis plots for $t_{ij} = 1$ (bold lines) and $t_{ij} = 2$ (dashed lines) as a function of: (d) α and A_{ij} when $\rho = 0.1$; (e) A_{ij} and ρ when $\alpha = 1$; (f) α and ρ when $A_{ij} = 1$.

4.2.3.2 Lambert distance and line-graph arrival time

The perspective to derive the Lambert distance is different from the line-graph arrival time. First, in the Lambert distance, it is assumed that the links between i and j are independent. The arrival time at j is simply given by the sum of the Lambert edge distance of all links that connect i and j. Second, we assume in the line graph that the arrival time of a node is influenced by its previous neighbour node, except the seed. This approach gives the n-th root in the Lambert function as our estimate.

Now, Let W_{ij} be the arrival time at node j as defined in (4.19). By multiplying the left-hand side with $e^{\alpha W_{ij}}$ and applying the arithmetic-geometric inequality, we

obtain

$$W_{ij}e^{\alpha W_{ij}} \geq \frac{n^n}{\rho^n \mathcal{T}(\Gamma_{ij}; n)} > \frac{n!}{\rho^n \mathcal{T}(\Gamma_{ij}; n)}, \tag{4.26}$$

or equivalently: $W_{ij} > t_{ij}$ for $n \geq 2$. The equality in (4.26) occurs when we consider a homogeneous line where all weights between two nodes have the same values.

4.2.3.3 Propagation of multiple seeds

We first derive the propagation of multiple seeds using a simple topology as illustrated in Figure 4.10. We assume that there are K seeds and each seed i_k is connected to

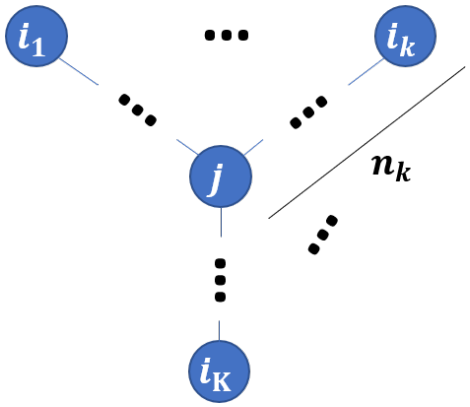


Figure 4.10: A simple configuration of multiple seeds.

the target j by a path $\Gamma_{i_k,j}$ with length n_k, $k = 1, ...K$. We can neglect all paths to node j when their starting nodes are not seeds. The initial value at each seeding node can vary $c_k(0) = \beta_k$, for $k = 1, ...K$, and zero otherwise. The ODE at node j is given by

$$\frac{\mathrm{d}c_j}{\mathrm{d}t} = \alpha c_j - \rho \sum_{k=1}^{K} L_{v_k,j} c_{v_k} \tag{4.27}$$

where v_k are the adjacent nodes of j and the ODEs at the other nodes are given by (4.23).

The solution for seeding and transient nodes follows (4.24). Substituting the solution of c_{v_k} to (4.27), we can write

$$\frac{\mathrm{d}c_j}{\mathrm{d}t} - \alpha c_j = \sum_{k=1}^{K} \rho^{n_k} \mathcal{T}(\Gamma_{i_k,j}; n_k) \frac{t^{n_k-1}}{n_k - 1!} \beta_k e^{\alpha t}. \tag{4.28}$$

Hence, the total concentration at node j is given by

$$c_j(t) = \sum_{k=1}^{K} \mathcal{T}(\Gamma_{i_k,j}; n_k) \frac{(\rho t)^{n_k}}{n_k!} \beta_k e^{\alpha t}. \tag{4.29}$$

We note that the concentration at node j is a sum of all partial contributions from multiple seeds. Using this feature, we construct an approximation of the total concentration and arrival times at node j.

We assume that a general initial condition $\mathbf{c}(0) = \mathbf{q}$ is given, with $0 \leq \max_i\{q_i\} < \mu \ll 1$, and consider a seed i_k. Following the analysis of the Lambert arrival times, we use (4.22) with $\beta = q_{i_k}$ to obtain

$$\tau_{i_k,j} = \frac{1}{\alpha} \log\left(\frac{(1-q_{i_k})\mu}{(1-\mu)q_{i_k}}\right) + W_{i_k,j}, \tag{4.30}$$

which estimates the arrival time at node j due to the seed at node i_k. Let us assume that the connectivity between the seed and target follows the two-node model where the weight approximation is $\tilde{A}_{i_k,j} = (\rho\tau_{i_k,j} \exp(\alpha\tau_{i_k,j}))^{-1}$. Therefore, we can construct an approximation of the contribution of the seed i_k to the concentration at target j by

$$\tilde{c}_j^{(i_k)}(t) = \mu \frac{t \exp(\alpha t)}{\tau_{i_k,j} \exp(\alpha\tau_{i_k,j})}. \tag{4.31}$$

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

$$\tilde{c}_j(t) = \sum_{k=1}^{K} \tilde{c}_j^{(i_k)}(t) = \mu t \, \exp(\alpha t) \sum_{k=1}^{K} \frac{1}{\tau_{i_k,j}} \, \exp(-\alpha\tau_{i_k,j}), \tag{4.32}$$

which reaches the level μ at time

$$\tilde{t}_j(\mu) = \frac{1}{\alpha} W_0\left(\frac{\alpha}{\sum_{i=1}^{N} \frac{1}{\tau_{ij}} \exp(-\alpha\tau_{ij})}\right). \tag{4.33}$$

Just as in the case of a single seed, the estimate (4.33) can be used as an initial guess to compute the linear arrival times for the case of multiple seeding locations.

4.2.3.4 Propagation of a single seed on multiple paths

We have derived an approximation of the total concentration at the target node due to multiple seeds. It turns out that the total concentration is just a sum of

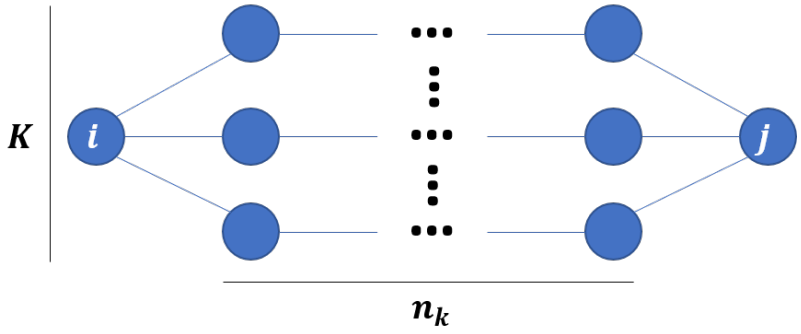


Figure 4.11: A simple configuration of multiple paths.

all contributions of multiple seeds. We now extend the notion of arrival times to incorporate multiple paths that connect a single seed and a target node.

We start with a simple case as illustrated in Figure 4.11. We assume that there are K distinct paths (i.e. each path consists of nodes registered in an unique path, except the seed and target node) that connect i to j where each path has length $n_k + 2$, for $k = 1, ..., K$. In one-dimensional line graph we see that the target node is influenced by its direct neighbour. However, it is possible that j has more than K neighbour nodes, e.g. M nodes, in this case. Hence, we have $M - K$ nodes that do not connect i and j and that can be neglected. We assume $c_i(0) = \beta$ and zero otherwise. Therefore, we have the governing system at node j given by

$$\frac{\mathrm{d}c_j}{\mathrm{d}t} = \alpha c_j - \rho \sum_{k=1}^{K} L_{v_k,j} c_{v_k} \tag{4.34}$$

where v_k are the nodes adjacent to j. The system at the other nodes is given by (4.23). This system is similar to (4.27).

Let $\Gamma_{i,j}^{(k)}$ be the k-th path that connects i and j with length $n_k + 2$. The solution of seeding and transient nodes follows (4.24). Substituting the solution of c_{v_k} to (4.34), we then can write

$$\frac{\mathrm{d}c_j}{\mathrm{d}t} - \alpha c_j = \sum_{k=1}^{K} \rho^{n_k+2} \mathcal{T}(\Gamma_{i,j}^{(k)}; n_k + 2) \frac{t^{n_k+1}}{n_k + 1!} \beta e^{\alpha t}. \tag{4.35}$$

Hence, the total concentration at node j is given by

$$c_j(t) = \sum_{k=1}^{K} \mathcal{T}(\Gamma_{i,j}^{(k)}; n_k + 2) \frac{(\rho t)^{n_k+2}}{n_k + 2!} \beta e^{\alpha t}. \tag{4.36}$$

We found that the total concentration at node j is just a sum of all partial contributions from the multiple paths like (4.29). Using this feature, we will construct an approximation of the total concentration and arrival times at node j.

We assume that a general initial condition $\mathbf{c}(0) = \mathbf{q}$ is given, with $0 \leq \max_i\{q_i\} < \mu \ll 1$, and consider the seed i. Let $W_{ij}^{\langle k\rangle}$ be the sum of the Lambert edge distances from i to j through the k-th path. Following the analysis of the Lambert arrival times, we use (4.22) with $\beta = q_i$ to obtain

$$\tau_{ij}^{\langle k\rangle} = \frac{1}{\alpha}\log\left(\frac{(1-q_i)\mu}{(1-\mu)q_i}\right) + W_{ij}^{\langle k\rangle}, \tag{4.37}$$

which estimates the arrival time at node j due to the seed at node i through the k-th path. Let us assume that the connectivity between the seed and target on each path follows the two-node model where the weight approximation is $\tilde{A}_{ij}^{\langle k\rangle} = \left(\rho\tau_{i,j}^{\langle k\rangle}\exp\left(\alpha\tau_{ij}^{\langle k\rangle}\right)\right)^{-1}$. Therefore, we can construct an approximation of the contribution of the seed i to the concentration at target j through the k-th path by

$$\tilde{c}_j^{\langle k\rangle}(t) = \mu\frac{t\exp(\alpha t)}{\tau_{ij}^{\langle k\rangle}\exp\left(\alpha\tau_{ij}^{\langle k\rangle}\right)}. \tag{4.38}$$

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

$$\tilde{c}_j(t) = \sum_{k=1}^{K}\tilde{c}_j^{\langle k\rangle}(t) = \mu t\,\exp(\alpha t)\sum_{k=1}^{K}\frac{1}{\tau_{ij}^{\langle k\rangle}}\,\exp(-\alpha\tau_{ij}^{\langle k\rangle}), \tag{4.39}$$

which reaches the level μ at time

$$\tilde{t}_j(\mu) = \frac{1}{\alpha}W_0\left(\frac{\alpha}{\sum_{i=1}^{N}\frac{1}{\tau_{ij}^{\langle k\rangle}}\exp(-\alpha\tau_{ij}^{\langle k\rangle})}\right). \tag{4.40}$$

We note that this form is similar to (4.33) and can use the estimate (4.40) as an initial guess to compute the linear arrival times when multiple paths are considered.

4.2.3.5 Many propagation schemes

In the case of multiple paths, we may find more complex propagation schemes than the previous simple one. When a single seed is considered, we could have a branching event when a single line splits into several paths, a narrowing event when multiple paths from the seed merge into a single line to the target, or there exists a bridge

or more between paths. We then investigate the propagation feature of these many propagation schemes.

Branching path. We first extend the case of multiple paths in Figure 4.11 into a *branching event* as illustrated in Figure 4.12. Any seed starts to flow along a line of length n then splits into K lines where each line ends up at node j and has length $n_k + 2$, $k = 1, ..., K$. The key to this type of propagation is the adjacent nodes of i_n

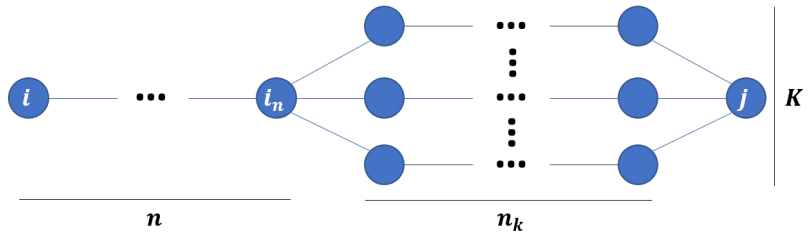


Figure 4.12: Branching case.

at each branch line. The concentration at these nodes is given by

$$c_{j_k}(t) = \rho A_{j_k,i_n} e^{\alpha t} \int_0^t c_{i_n}(s) e^{-\alpha s} \mathrm{d}s, \tag{4.41}$$

where j_k are the adjacent nodes of i_n, at the branches, and c_{i_n} is given by (4.24).

Following the previous analysis, we use (4.36) and the concentration at j is given by

$$c_j(t) = \sum_{k=1}^{K} \rho^{n_k+2} \mathcal{T}(\Gamma_{i_n,j}^{\langle k \rangle}; n_k + 2) e^{\alpha t} \left[\int_{\mathcal{R}} c_{i_n}(s_0) e^{-\alpha s_0} \mathrm{d}\mathcal{S} \right], \tag{4.42}$$

where $\mathcal{R} = \{(s_0, s_1, ...s_{n_k+2}) | s_0 \leq s_1, 0 \leq s_1 \leq s_2, ..., 0 \leq s_{n_k+2} \leq t\}$ and $\mathrm{d}\mathcal{S} = \mathrm{d}s_0 \mathrm{d}s_1 ... \mathrm{d}s_{n_k+2}$. Then, the concentration at j is given by

$$c_j(t) = \sum_{k=1}^{K} \mathcal{T}(\Gamma^{\langle k \rangle}; n_k + 2) \mathcal{T}(\Gamma_{i,i_n}; n) \frac{(\rho t)^{n+n_k+2}}{n + n_k + 2!} \beta e^{\alpha t}. \tag{4.43}$$

We note that (4.43) is a sum of the contributions over K paths as similar to (4.36). One then can derive the approximation of concentrations and arrival times following (4.39) and (4.40), respectively.

Narrowing path. Next, we consider a system when multiple paths are focusing on a single line that ends up at j. We call this event the *narrowing path*. In this case, let there are K paths where each path has length $n_k + 2$ and they meet in one node j_K connected to j along a line graph with length n as illustrated in Figure 4.13).

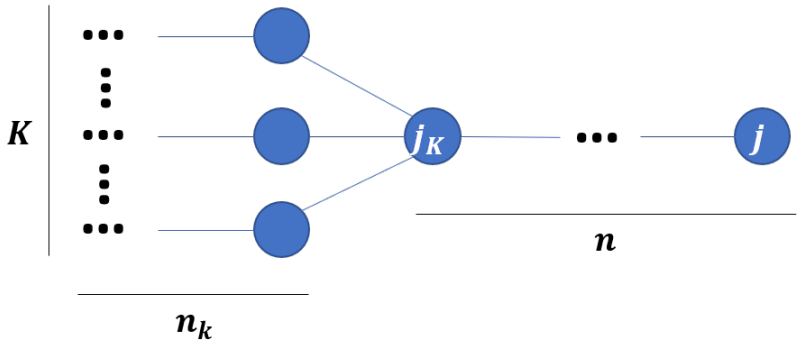


Figure 4.13: Narrowing case.

The concentration at j is to depend on concentration at j_K and is given by

$$c_j(t) = \rho^n \mathcal{T}(\Gamma_{j_K,j}; n) e^{\alpha t} \left[\int_{\mathcal{R}} c_{j_K}(s_0) e^{-\alpha s_0} \mathrm{d}\mathcal{S} \right], \tag{4.44}$$

where $\mathcal{R} = \{(s_0, s_1, ...s_{n_k+2}) | 0 \leq s_0 \leq s_1, 0 \leq s_1 \leq s_2, ..., 0 \leq s_{n_k+2} \leq t\}$ and $\mathrm{d}\mathcal{S} = \mathrm{d}s_0 \mathrm{d}s_1 ... \mathrm{d}s_{n_k+2}$. The concentration at j_K is given by the sum of all paths connected to a seeding node or more. Therefore, c_{j_K} follows (4.36), (4.29), or superposition of both. Without loss of generality, we assume that the first K_1 paths connect j_K to a seeding node i_1 and the remaining paths are connected to a distinct seed i_2. Then, the concentration at j is given by,

$$\begin{aligned} c_j(t) = &\sum_{k=1}^{K_1} \mathcal{T}(\Gamma_{j_K,j}; n) \mathcal{T}(\Gamma^{\langle k \rangle}_{i_1,j_K}; n_k + 2) \frac{(\rho t)^{n+n_k+2}}{n + n_k + 2!} \beta e^{\alpha t} \\ &+ \sum_{k=K_1+1}^{K} \mathcal{T}(\Gamma_{j_K,j}; n) \mathcal{T}(\Gamma^{\langle k \rangle}_{i_2,j_K}; n_k + 2) \frac{(\rho t)^{n+n_k+2}}{n + n_k + 2!} \beta_k e^{\alpha t}. \end{aligned} \tag{4.45}$$

We note that the solution is also a sum of the contributions over K paths as similar to (4.43). One can derive the approximation of concentrations and arrival times following (4.39) and (4.40), respectively.

Bridge between paths. Another important case in a network is the possibility of having bridges between paths as it creates additional paths. We now consider a simple case where two paths are connected with an edge as illustrated in Figure 4.14.

We consider two distinct paths of length n_1 and n_2, which do not have the same vertices except for the endpoint. Without loss of generality, we assume that we have multiple seeding nodes, i.e. i_1 and i_2. The two paths are connected through an edge from node k_1 to k_2. In the absence of this edge, concentration at these nodes is given by (4.24) and the concentration at j is given by (4.29). When this edge is introduced,

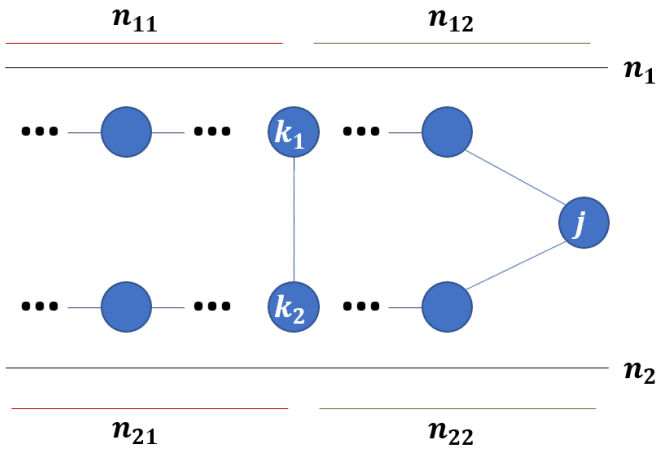


Figure 4.14: Communicating case.

we now have two additional routes. Considering all possible paths through node k_1 and k_2, the concentration at j is

$$\begin{aligned} c_j(t) &= \rho^{n_{12}+1}\mathcal{T}(\Gamma_{k_1,j};n_{12})A_{k_1,k_2}e^{\alpha t}\left[\int_{\mathcal{R}} c_{k_2}(s_0)e^{-\alpha s_0}\mathrm{d}\mathcal{S}\right] \\ &+ \rho^{n_{22}+1}\mathcal{T}(\Gamma_{k_2,j};n_{22})A_{k_1,k_2}e^{\alpha t}\left[\int_{\mathcal{R}} c_{k_1}(s_0)e^{-\alpha s_0}\mathrm{d}\mathcal{S}\right] \\ &+ \rho^{n_{12}}\mathcal{T}(\Gamma_{k_1,j};n_{12})e^{\alpha t}\left[\int_{\mathcal{R}} c_{k_1}(s_0)e^{-\alpha s_0}\mathrm{d}\mathcal{S}\right] \\ &+ \rho^{n_{22}}\mathcal{T}(\Gamma_{k_2,j};n_{22})e^{\alpha t}\left[\int_{\mathcal{R}} c_{k_2}(s_0)e^{-\alpha s_0}\mathrm{d}\mathcal{S}\right]. \end{aligned} \tag{4.46}$$

Since we already separate the paths when the bridge is considered, we assume that the concentration at k_1 depends on the solution of the line graph connecting the seed i_1 to k_1 only. This assumption is also imposed to k_2 where it is connected to i_2. Let the initial condition be $c_{i_1}(0) = \beta_1$, $c_{i_2}(0) = \beta_2$, and zero otherwise. Hence, the concentration at j is given by,

$$\begin{aligned} c_j(t) &= \mathcal{T}(\Gamma_{i_2,k_2};n_{21})A_{k_1,k_2}\mathcal{T}(\Gamma_{k_1,j};n_{12})\frac{(\rho t)^{n_{21}+n_{12}+1}}{n_{21}+n_{12}+1!}\beta_1 e^{\alpha t} \\ &+ \mathcal{T}(\Gamma_{i_1,j};n_1)\frac{(\rho t)^{n_1}}{n_1!}\beta_1 e^{\alpha t} \\ &+ \mathcal{T}(\Gamma_{k_1,j};n_{12})A_{k_1,k_2}\mathcal{T}(\Gamma_{i_2,k_2};n_{21})\frac{(\rho t)^{n_{12}+n_{21}+1}}{n_{12}+n_{21}+1!}\beta_2 e^{\alpha t} \\ &+ \mathcal{T}(\Gamma_{i_2,j};n_2)\frac{(\rho t)^{n_2}}{n_2!}\beta_2 e^{\alpha t}. \end{aligned} \tag{4.47}$$

$\Gamma_{i_1,j}$ and $\Gamma_{i_2,j}$ are lines that connect the seeds with j without considering the bridge from k_1 to k_2. We note that new paths are created once a bridge is added. Similar to

previous schemes, the concentration depends on all possible paths that connect the seeds and target.

In this scheme, the total number of paths increases following the binomial rule as more bridges between two distinct paths are introduced. Indeed, if there are n bridges between two paths, the total number of paths is 2^{n+1}.

We have investigated the propagation of proteins on a network for multiple paths and multiple seeds. First, the concentration at a target node always considers all possible paths from multiple paths or multiple seeds. It simply takes the sum of all paths in that network configuration. Second, when multiple paths and multiple seeds appear in a network, we just take the sum of all paths. The feature suggests that each path adds its contribution to the total concentration. Third, we can find an approximation of the total concentration by following the process of (4.32) and (4.39). Then, we can also determine the approximation of arrival times which follows (4.33) or (4.40). As an additional remark, one needs to include multiple paths when computing the arrival times from an unweighted graph. We observe (although it is not shown) that if we include more shortest paths in the computation, the accuracy of estimates increases. This analysis allows us to identify additional path contribution to the total concentration, and ultimately, to the arrival times.

4.2.4 Nonlinear estimate

We have obtained estimates for the arrival times mostly from a linear approach and neglected the nonlinear term in the original system. Despite their merit, these methods do not take into account the coupling between the nonlinear sigmoid growth and the diffusion among nearby nodes. A third approach is to include the nonlinear sigmoid and a small parameter that naturally appears in the growth-dominated case.

We 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. Instead of expanding all solutions, we consider an alternative asymptotic method that captures important features of the solutions and improves the computation of the arrival times.

We consider the canonical case to be our ansatz and start with the solution with $\varepsilon = 0$ given by,

$$\tilde{c}_i = \frac{K_i e^{\alpha t}}{1 - K_i \left(e^{\alpha t} - 1\right)}, \quad i = 1, \ldots, N. \tag{4.48}$$

We use a nonlinear version of the method of variation of constants instead of determining the arbitrary constant from the initial conditions. We assume that this constant is allowed to vary in time at higher order in ε,

$$K_i = A_i + \varepsilon^{\gamma_i} B_i(t), \tag{4.49}$$

where γ_i is a natural number that defines the shortest path length between node i and the seed. The value is not a Lambert distance between node i and the seed. Rather, it is the number of hops between node i and the seed with respect to the Lambert metric, as previously defined. With this ansatz, we can solve iteratively for the unknown B_i for a subsequent node with increasing distance from the seeding node.

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

$$\tilde{c}_s = \frac{\beta e^{\alpha t}}{1 - \beta\left(e^{\alpha t} - 1\right)}, \tag{4.50}$$

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 equation for B_i is

$$\frac{\mathrm{d}B_i}{\mathrm{d}t} = -L_{is} e^{-\alpha t} \tilde{c}_s, \tag{4.51}$$

that can be easily integrated to obtain

$$B_i(t) = \frac{\beta L_{is}}{\alpha(\beta - 1)} \left(\alpha t - \log\left(\beta\left(e^{\alpha t} - 1\right) + 1\right)\right). \tag{4.52}$$

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

$$M_{ij} = \begin{cases} L_{ij} & \text{if } \gamma_i - \gamma_j = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{4.53}$$

and the variables

$$Z_i = \begin{cases} B_i, & \text{if } i \neq s \\ e^{-\alpha t}\tilde{c}_s & \text{if } i = s, \end{cases} \tag{4.54}$$

then, these new variables are solutions of the system

$$\frac{\mathrm{d}Z_i}{\mathrm{d}t} = \sum_{j=1}^{N} M_{ij} Z_j, \qquad Z_i(0) = 0, \qquad i \neq s. \tag{4.55}$$

Remarkably, the solution of this linear system can be obtained explicitly in terms of the polylogarithm function Li_d where d is the distance to the seeding node. Once substituted in (4.48), these *asymptotic solutions* satisfy both the initial conditions and the asymptotic behavior for a long time. Figure 4.15 shows these asymptotic solutions for our 5-node network together with the computations of the arrival times.

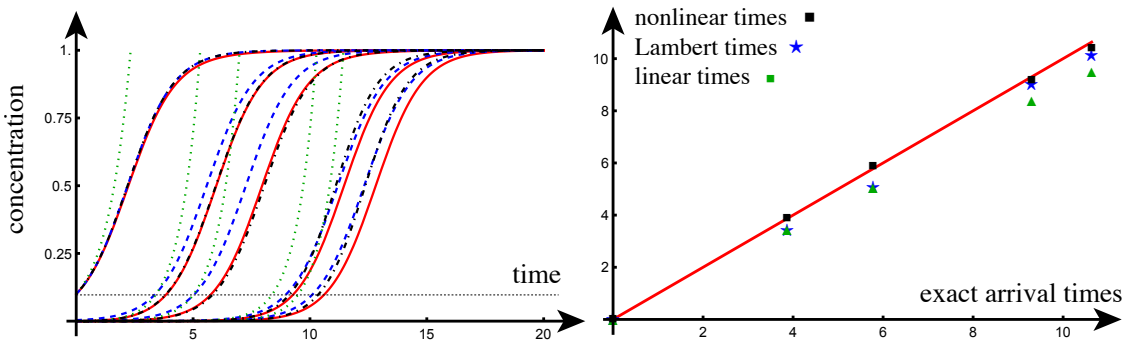


Figure 4.15: a. Exact (red), Lambert (blue, dashed), linear (green, dotted), and asymptotic (black, dash-dotted) solutions with $\alpha = 1, \rho = 1/100, p_1(0) = \mu = 1/10$. b. Asymptotic Lambert arrival times (blue stars), compared to linear arrival times (green squares).

A few comments regarding the nonlinear estimate are as follows. The method improves the Lambert solution. It is because the solutions naturally take into account the contributions from all paths of length d rather than just a single path. Although the solutions contain paths with the same length, 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. The handicap of this method is its long expressions. In addition, this approach may fail to generalise to other problems due to its particular ansatz.

4.3 Systematic comparison

In this section, we test our methods to see their robustness through error computation. In addition, we check the variability of mean arrival times generated from many connectomes.

4.3.1 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

$$\mathcal{E}_{\text{approx}} = \frac{\sum_{i=1}^{N} |\widehat{\tau}_i - \tau_i|}{\sum_{i=1}^{N} \tau_i}, \qquad \text{approx} \in \{\text{linear, Lambert}\}. \tag{4.56}$$

We fix the growth constant $\alpha = 1/2$ and systematically test our approach on the brain connectome by varying the weight of the edge connection, the size of the network, the diffusion constant ρ, and the initial condition. We also choose a few different one-parameter networks describing different topologies despite our main interest in the connectome. We use different network sizes for a systematic evaluation of the approximation errors. The connectomes have edges with weights A_{ij} between nodes i and j, defined as the ratio of the number of axonal fibers n_{ij} between the nodes and the mean fiber length ℓ_{ij} power [144],

$$A_{ij} = \frac{n_{ij}}{\ell_{ij}^{k}}, \quad i, j = 1, \ldots, N, \tag{4.57}$$

with $k = 0$, $k = 1$, and $k = 2$ respectively define the *length-free*, the *ballistic*, and the *diffusive* weights. Particular to the diffusive weight, the length is squared in order to capture the main scaling properties of diffusion. This type of weight becomes our primary choice to perform the application of our approaches to neurodegenerative disease in the next section.

4.3.2 Analysis on the weights and the network size

4.3.2.1 Varying the weights

To understand the possible dependence of our method on the choice of edge weights, we consider the one-parameter family of adjacency matrices (4.57) for the $N = 83$ connectome where we take $k \in [0, 2]$. These matrices cover all the choices previously made in the literature for the propagation of toxic proteins on the structural connectome ($k \in \{0, 1, 2\}$). We see in Figure 4.16a that the errors from both the Lambert and the linear methods remain small and do not depend on the choice of weights).

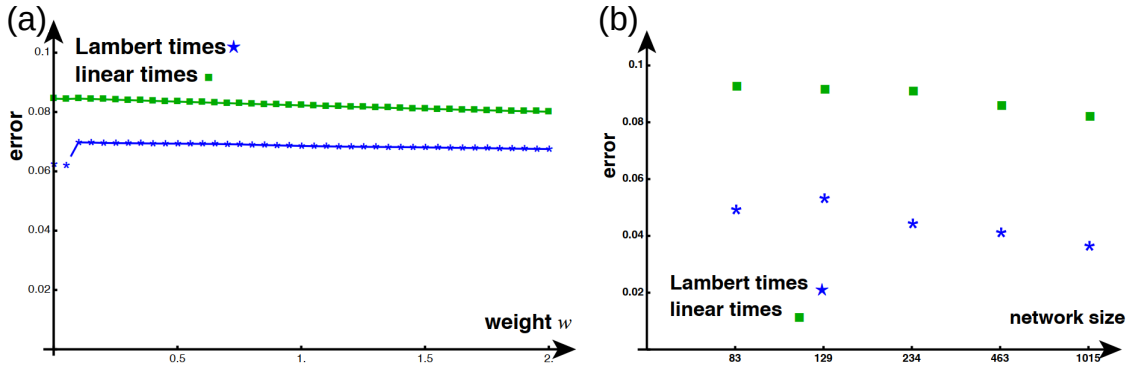


Figure 4.16: Error analysis with respect to (a) changes in weight and (b) changes in network sizes. Parameters for both: $\alpha = 0.5$/year and $\rho = 0.01$/year; initial conditions: $c_{27}(0) = c_{68}(0) = 1/20$ and $c_i = 0$ for all other $i \in \{1, \ldots, N\}$). We observe that the method does not depend strongly on either the weight or the size (as long as the system remains in the growth-dominated regime).

4.3.2.2 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. Figure 4.16b demonstrates the results of the five runs. The error analysis shows that there is little effect associated with changes in network size.

4.3.3 Analysis on the diffusion constant and the initial conditions

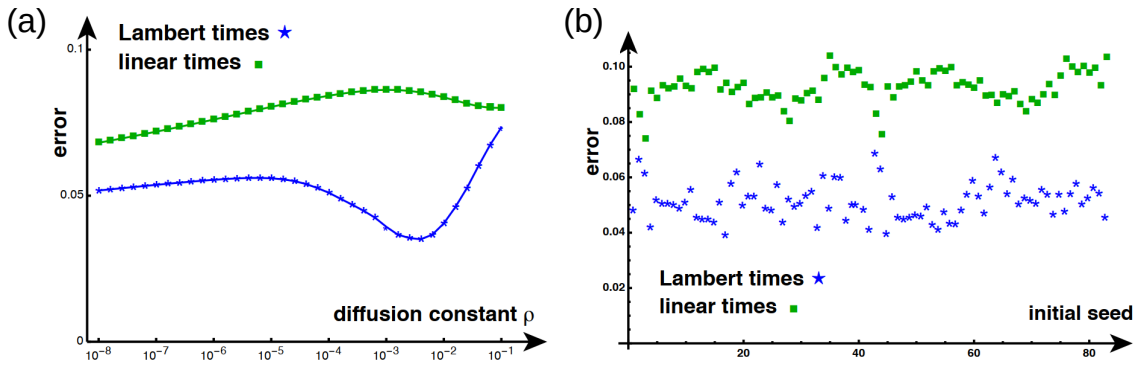


Figure 4.17: Error analysis with respect to changes in (a) changes in diffusion ($k = 2$ and $\rho = 0.01$/year) and (b) changes in initial conditions. Parameter for both: $\alpha = 0.5$/year. Initial conditions: $c_i^{(i)} = 1/10$ and $c_j^{(i)} = 0$ for all other $j \in \{1, \ldots, 83\}$.

4.3.3.1 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: $c_{27}(0) = c_{68}(0) = 1/20$. We observe in Figure 4.17a that the error remains small and does not strongly depend on the value of ρ. As expected when ρ is of order one, the Lambert estimates break down since we are not in the growth-dominated regime anymore.

4.3.3.2 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 non-zero concentration at a single-node i and all other initial conditions to zero. Figure 4.17b shows that there is little effect associated with choices in initial conditions. The typical error associated with the Lambert method is about 1/2 to 3/4 the error associated with the linear method.

4.3.4 Varying the topology

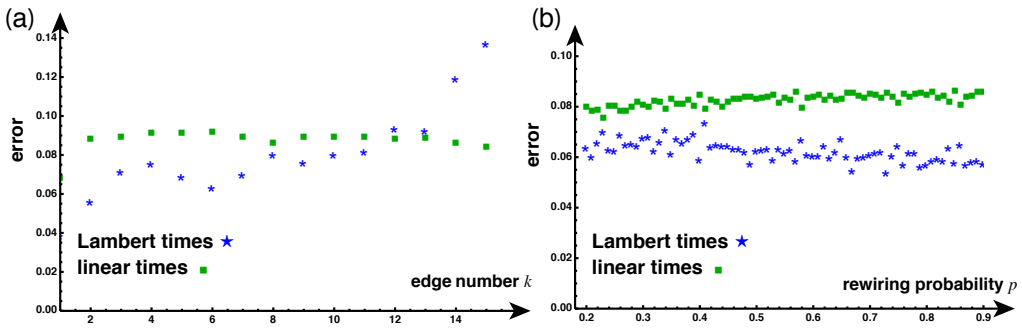


Figure 4.18: Error analysis with respect to changes in topology. (a) Scale-free network where k edges are added at each step. (b) Small-world network (with average degree of node 10) with rewiring probability p. Parameter: $N = 500$, $\alpha = 1/\text{year}$, $\rho = 0.001/\text{year}$. Initial conditions: $c_1(0) = 1/10$ and $c_i = 0$ for $i = 2, \ldots, N$.

To understand the possible dependence of our method on the topology of the network, we use different classes of network. We consider two types of graphs: (a) a distribution of scale-free network where a graph is built by adding a new vertex with k edges at each step; and (b) a distribution of small-world networks with rewiring probability p. We observe that, for scale-free network, the linear provides good estimates while the Lambert method is not as reliable as we thought. This is due to the

fact that multiple shortest paths are possible and the method only considers a single one of these. We also note that the linear method is robust and that the error does not depend on the topology.

4.3.5 Variability

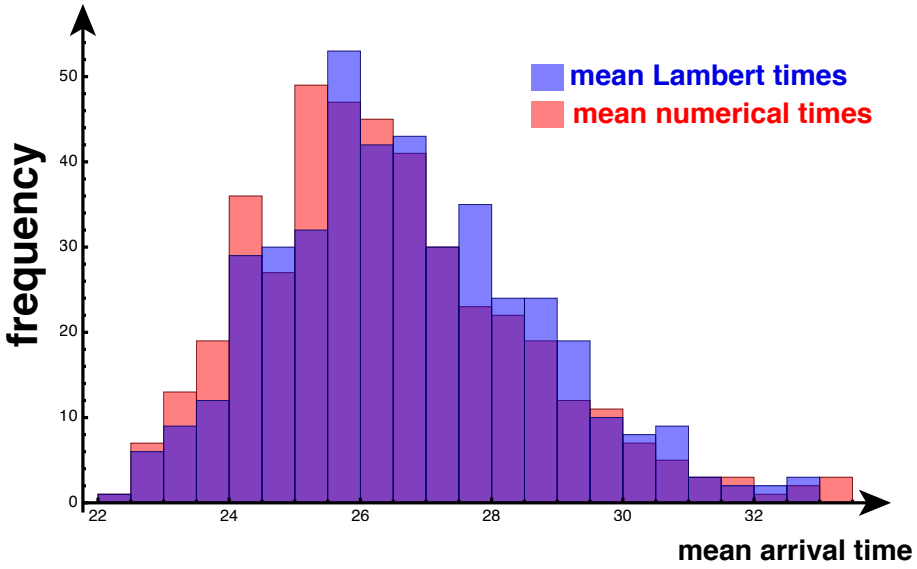


Figure 4.19: Histogram of the mean arrival times for each of the 426 structural connectomes in the reference data set obtained either from the exact (numerical) method or its Lambert approximation.

It is of interest to see if our methods can capture the variability between different individuals. We consider the 426 individual connectomes included in the Budapest data set and extract for each one a diffusion-weighted symmetric adjacency matrix. Since most of these connectomes are disconnected, we superimpose to it a small perturbation (taken to be 1/100) given by the average connectome and normalise the resulting adjacency matrix so that its maximum element is one.

We compute, for each connectome, both the exact (numerical) mean arrival time and its Lambert approximation for the same values of the parameters and initial conditions used in the main text ($N = 83, \alpha = 0.5$/year, $\rho = 0.01$/year. $c_{27}(0) = c_{68}(0) = 1/20$ and $c_i = 0$ for all other $i \in \{1, \ldots, N\}$). The distribution of arrival times is given in Fig. 4.19 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 compared to 2.0 for the exact method). We conclude that the Lambert method captures both the mean and variability found in individual brain connectome.

4.4 Application to neurodegenerative disease

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

4.4.1 Arrival times and Braak staging

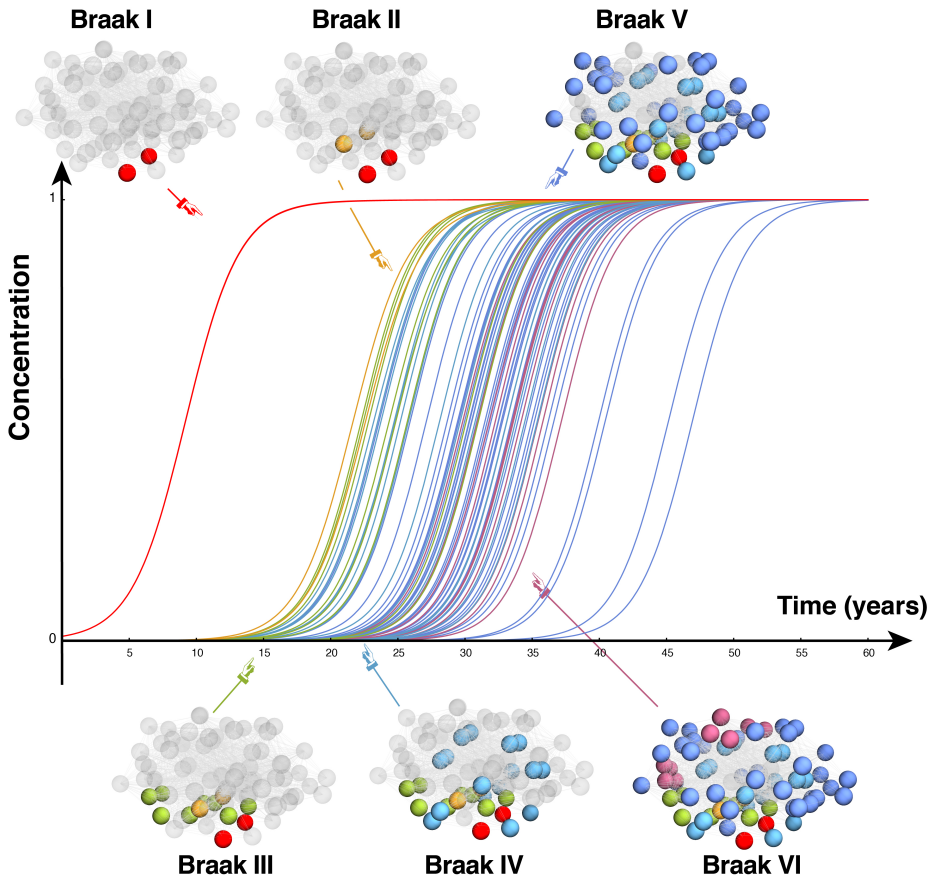


Figure 4.20: Starting at the entorhinal cortex, the dynamics of the network Fisher-KPP equations recovers some of the key aspects of Braak staging $N = 83, \alpha = 0.5$/year, $\rho = 0.01$/year. Initial conditions: $C_{27}(0) = c_{68}(0) = 1/100$ and $c_i = 0$ for all other $i \in \{1, \ldots, N\}$. 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.

Despite the overall simplicity of the network-organised Fisher-KPP model, its dynamics captures the generalised progression of τP neurofibrillary tangles (NFTs) observed in post-mortem studies. In particular, histopathological studies have noted that NFT follow a six-stage sequence [17, 18], called the *six Braak stages* [44]. In this work, we study the progression in the six Braak regions used in the standardised positron emission tomography AV-1451 data pipeline [2] of the Alzheimer's Disease

Neuroimaging Initiative. Figure 4.20 shows the dynamics of (4.1) using the lowest resolution $N = 83$ connectome with example parameters in the growth-dominated regime. In this case, the entorhinal seeding sites are nodes 27 and 68.

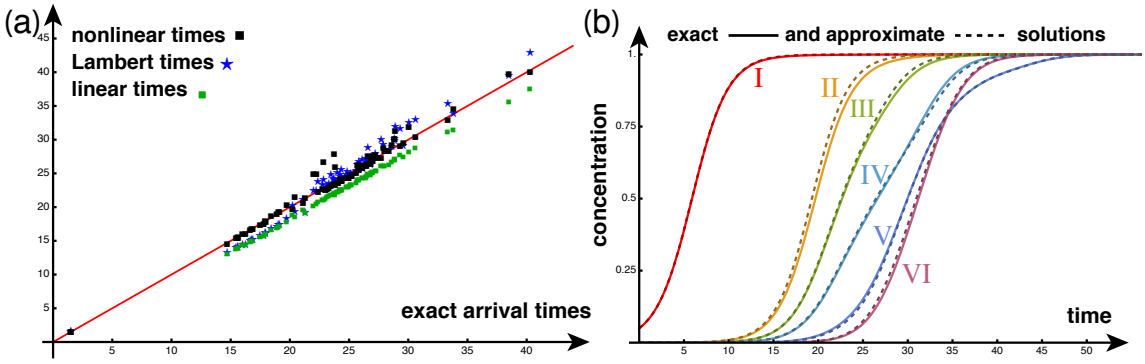


Figure 4.21: Arrival times at $\mu = 1/10$ for $N = 83$ connectome (left) and average concentration (right) in each Braak region (parameters and color coding coincide with Fig. 4.20). Solid curves are the numerical solution and the dashed curves their approximation from the asymptotic expansion. Initial conditions chosen so that the total concentration in the entorhinal cortex is 1/10: $c_{27}(0) = c_{68}(0) = 1/20$ and $c_i = 0$ for all other $i \in \{1, \ldots, N\}$.

We apply the three methods for arrival times on this problem for the same parameters as in Figure 4.20 and compute the approximate solutions. Figure 4.21 shows the three arrival times and compares, for each Braak region, the exact numerical solution (solid curves), with the asymptotic solutions (dashed curves). To visualise the results and the quality of the approximations, we average the concentration for each Braak region (total concentration over nodes in that Braak region divided by the number of nodes). We noted that the asymptotic solution is excellent on the entire domain and, remarkably, correctly captures the staging pattern observed in Alzheimer's disease. We also compare the relative error following (4.56) and the scores are $\mathcal{E}_{\text{linear}} \approx 84 \times 10^{-3}$, $\mathcal{E}_{\text{Lambert}} \approx 41 \times 10^{-3}$, and $\mathcal{E}_{\text{nonlinear}} \approx 24 \times 10^{-3}$.

4.4.2 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 the previous chapter or Putra et al [144]. The idea is to only consider the ordering of the arrival times rather than their numerical values.

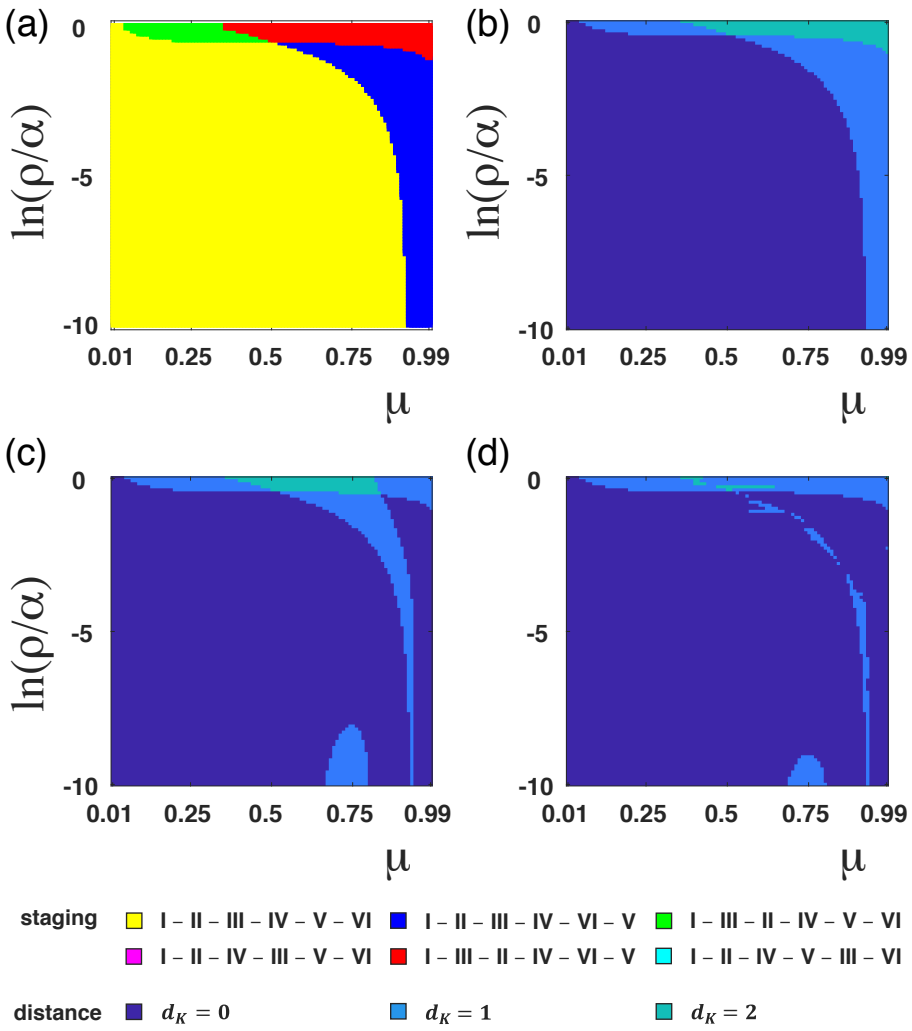


Figure 4.22: (a) An exact braid surface computed using 4.1 on the $N = 83$ brain connectome. Kendall tau errors for the approximated braid surfaces using: (b) the linear method, (c) the Lambert method, and (d) the nonlinear method. An initial seeding is chosen so that the average concentration in the entorhinal cortex is less than 1/100: $c_{27(0)} = c_{68}(0) = 1/200$ and $c_i = 0$ for all other $i \in 1, \ldots, N$. Yellow regions correspond to a canonical connectome Braak staging.

Following [144], we can plot stagings, across all possible values of the threshold μ, by using a braid diagram and a braid surface that summarises a parametric family of braid diagrams. To test the use of our methods, we use the *Kendall tau* distance [100] to measure the distance between two stagings as the minimum number of transpositions necessary to transform one staging into the other. For example, the Kendall tau distance between staging $(1, 3, 2)$ and $(1, 2, 3)$ is $d_K((1, 3, 2), (1, 2, 3)) = 1$, since transposing the last two values of $(1, 3, 2)$ brings the stagings into agreement. Using

this notion of distance, we can evaluate the errors between a braid surface from 4.1 and those from the linear, Lambert and nonlinear methods.

Figure 4.22a 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 [17, 18]. Figures 4.22b–d show the Kendall Tau distance between the exact Braid surface and the Braid surfaces constructed using the three arrival time estimates.

We note some important features regarding the comparison. First, the errors generally emerge 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. This feature is quite obvious since the methods have been designed under the general assumptions that the threshold is small and the dynamics is growth-dominated. Second, the nonlinear approximation (Figure 4.22d) has the lowest distribution of error, followed by the Lambert distance (Figure 4.22c) and the linear estimate (Figure 4.22b). This shows a good agreement of our method as shown in the previous example Figure 4.21. Overall, we see that all methods are excellent in recovering the exact staging in the growth-dominated regime and can therefore be used systematically for that purpose. These general observations again hold true for the high resolution $N = 1015$ connectome (not shown).

4.5 Concluding remark

In this chapter, we have presented methods to address the arrival-time problem that occurs in the modelling of toxic protein propagation throughout the brain network. Our objective is not to just obtain methods or to compute the arrival times. It is more important to extract estimates from a system coupling a nonlinear process with the diffusion given by the graph Laplacian in order to understand its dynamics in a meaningful way. We derived three estimates.

The first one, the linear estimate, is obtained by linearising the full system around the healthy state. We initially obtained the solution for toxic protein concentration at each node. We then solved the transcendental equation and obtain the arrival time of each node at a certain threshold. We noted that the linear method can be used for any system.

The second approach leads to the Lambert estimate. We initially computed the arrival times between two neighbouring nodes. Since these arrival times are independent of other connections, the estimate is simply a sum of arrival times within

a shortest path between two nodes. Remarkably, it satisfies the metric properties. We noted that the Lambert estimate does not depend on the choice of initial values, however, is strongly affected by the topology of networks.

The third approach, the nonlinear estimate, is an improvement on the Lambert solution obtained by considering a natural small parameter. In this method, we included nonlinearity in the solution and were able to take into account the multiple paths with the same length within the solution. Although we have to compute the solution iteratively from the seeding node to the target node in a shortest path with respect to the Lambert metric, the solution can be represented as a sum of polylogarithm functions. However, the method strongly depends on the exact solution for the single-node problem.

These methods allowed us to explore the propagation process from a seed to its neighbouring nodes. From the Lambert estimate and the nonlinear solution, they showed that the first phase of invasion depends on the shortest path with respect to the Lambert metric. Although the topology and the weights of the network strongly affect the time of invasion, these methods can accommodate any topology of networks under the Fisher-Kolmogorov-Petrovsky-Piskunov dynamical process. Ultimately, the arrival time estimates can help us understand the complex coupled process from an analytical perspective.

Chapter 5

Heterogeneous Systems and Application to Neurodegenerative Disease

5.1 Overview

5.1.1 Regional heterogeneity

Cognitive decline due to neurodegeneration is associated with specific brain structural alterations and neuronal loss. Common neurodegenerative diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), and Amyotrophic Lateral Sclerosis (ALS) show a feature of regional neurodegeneration [95]. In AD particularly, some regions are affected severely whilst leaving others relatively unimpaired [17]. Many empirical studies have shown the varying spatiotemporal distribution of biomarkers, for example, τP propagation throughout the brain. This differentiation indicates that some brain regions are selectively prone to the cause of neurodegeneration [132]. However, the mechanism causing different brain regions' variational neurodegeneration remains elusive.

The progression of neurodegeneration is influenced by key factors such as toxic protein growth and spreading throughout the brain network. The disease starts in a seeding region and spreads to neighboring regions before invading all connected regions [18, 23, 95, 188]. Subtle differences between brain regions as the disease develops are believed due to substantial cellular composition in different brain regions [132]. Regional susceptibility may suggest heterogeneous local growth, which is a factor that can create differences in disease propagation between regions [198]. Some brain regions are hotspots for the disease while others have almost no infections. The existence of hotspots means that faster-growing regions can influence growth rates on a

larger scale [29].

From a mathematical perspective, regional vulnerability features can be included in a mathematical model as heterogeneous parameters representing local characteristics. Introducing local growth suggests faster disease propagation and possibly shorter transmission paths [5]. The main question is to understand the resulting dynamics of different heterogeneous systems. For instance, if two regions have slight local differences, do they show significantly distinct dynamics? In terms of propagation dynamics, how long does it take for a process such as a disease to develop into a full-scale invasion when it starts at a seed location in the presence of heterogeneous features? If the growth rate changes slightly, do we observe a drastic change in dynamics? How does the change in growth rates affect overall disease propagation? Despite the complexity of heterogeneous systems, the question is to reveal the dynamics due to heterogeneity in a system and derive related estimates rather than just applying the model to available data. In this study, we propose exploring the effect of local growth on physical transport dynamics on a network.

5.1.2 Exploration on heterogeneous systems

In this study, we restrict our exploration to linear coupling dynamics between nodes, which resembles the graph Laplacian used for global disease progression [23, 57, 114] and the modelling of neurodegenerative diseases [120]. Our study is a mathematical exploration of the dynamical system and notable measures such as *the arrival time problem*, which measures the time it takes for a quantity of interest to reach a determined threshold at each node, and *the staging problem*, which determines the sequence of regional spreading throughout the brain at different thresholds and parameter values.

We consider a general spatiotemporal system with a form

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{f}(t, \mathbf{p}; \boldsymbol{\Theta}, \boldsymbol{\Psi}_\Omega), \tag{5.1a}$$

$$\mathbf{p}(0) = \mathbf{q}, \tag{5.1b}$$

where $\mathbf{p}(t) = (p_1, ..., p_N)$ denotes a quantity, for example, a protein concentration, evolving on a connected and undirected network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a collection of N nodes $\mathcal{V}$, representing regions of interest, and edge set $\mathcal{E}$, representing connections between these regions. The quantity $p_i(t)$ corresponds to the observed concentration in node $v_i \in \mathcal{V}$ and evolves according to dynamics (5.1). Let $\mathbf{q} = (q_1, ..., q_N)$ and $q_i = p_i(0)$ is the initial concentration at that node. The quantity $\boldsymbol{\Theta}$ represents the global

parameters of the system (5.1) and $\mathbf{\Psi}_\Omega$ represents regional parameters associated with node $v_i \in \mathcal{V}$. Hence, the dynamics in (5.1) needs the regional information from each node of network $\mathcal{V}$. We further assume that the dynamics of the system is such that starting from an initial condition, where all concentrations are taken to vanish except in a few nodes, and the system will evolve asymptotically to a state where all concentrations reach their maximal value.

We recall the canonical arrival-time problem, in which 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 $\tau_i = \tau(\mu)$ such that $c_i(\tau_i) = \mu$ (with $\beta \leq \mu < 1$ so that $\tau_i \geq 0$ and finite with the equality $\tau_i = 0$ attained if only if $i = s$ and $\mu = \beta$).

Next, we recall the staging problem. Let Ω_j for $j = 1, ..., J$ be a nonoverlapping collection of nodes, such that $\Omega_j \subseteq \mathcal{V}$ and $\Omega_j \cap \Omega_k =$ if $j \neq k$. Let $T \in [0, 1]$ be an arbitrary threshold value. As the concentration evolves according to (5.1), we define the weighted average concentration P_j

$$P_j = \frac{1}{\sigma_{\Omega_j}} \sum_{i \in \Omega_j} p_i, \quad j = 1, ..., J,$$

where

$$\sigma_{\Omega_j} = \sum_{i \in \Omega_j} p_{i,\infty},$$

with $p_{i,\infty}$ represents the nontrivial asymptotic concentration at node i. We record the time when P_j first reaches the threshold T, $P_j(t_j) = T$. This process produces an ordering of the regions Ω_j according to the ordered arrival times called an observed staging pattern $\{\Omega_{j_1}, \Omega_{j_2}, ..., \Omega_{j_J}\}$. The generalised staging problem is to probe observable staging patterns subject to varying $\mathbf{\Theta}$ for a given $\mathbf{\Psi}_\Omega$.

5.2 Mathematical models for heterogeneous system

In this section, we discuss the modelling of prion-like protein propagation with heterogeneous parameters. We consider unequal carrying capacities or growth rates in the system.

5.2.1 Heterogeneous carrying capacities

We start with a coupled model of healthy and toxic protein following the heterodimer kinetics [142] from which we aim to derive a model of toxic protein propagation that

includes regional information. We introduce the total amount of healthy and toxic protein at node i at time t as $p_i = p_i(t)$ and $\tilde{p}_i = \tilde{p}_i(t)$, respectively. τP PET scans show that cognitively normal patients also develop toxic protein molecules in their brains. Therefore, we assume that there is a minimum concentration of toxic proteins such that $\tilde{p}_i \neq 0$ at the healthy state of the Heterodimer dynamics. We model the dynamics of healthy and toxic proteins

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = -\rho \sum_{j=1}^{N} L_{ij} p_i + k_0 - k_1 p_i - k_{12} p_i (\tilde{p}_i - \tilde{p}_{i,0}), \quad i = 1, ..., N, \tag{5.2}$$

$$\frac{\mathrm{d}\tilde{p}_i}{\mathrm{d}t} = -\rho \sum_{j=1}^{N} L_{ij} (\tilde{p}_i - \tilde{p}_{i,0}) - \tilde{k}_1 (\tilde{p}_i - \tilde{p}_{i,0}) + k_{12} p_i (\tilde{p}_i - \tilde{p}_{i,0}), \quad i = 1, ..., N. \tag{5.3}$$

We use the standard graph Laplacian $\mathbf{L} = [L_{ij}]$ given by

$$\mathbf{L} = \mathbf{D} - \mathbf{A}, \tag{5.4}$$

where $\mathbf{A} = [a_{ij}]$ represents the adjacency $N \times N$ matrix of an undirected network $\mathcal{G}$ and $\mathbf{D}$ is the degree matrix, a diagonal matrix with entries $D_{ii} = d_i = \sum_j a_{ij}$ for $i = 1, ..., N$. Close to the initial healthy state, $p_i \gg \tilde{p}_i$, the system (5.2) can be approximated by a Fisher-KPP equation. The condition of $p_i \gg \tilde{p}_i$ and its close-to-uniform distribution across the network imply that $\mathrm{d}p_i/\mathrm{d}t \approx 0$ and $\sum_j L_{ij} p_j \approx 0$. With these assumptions, the amount of healthy protein in each node can be approximated by

$$k_0 - k_1 p_i - k_{12} p_i (\tilde{p}_i - \tilde{p}_{i,0}) = 0, \text{ thus } p_i = \frac{k_0}{k_1 + k_{12}(\tilde{p}_i - \tilde{p}_{i,0})}, \tag{5.5}$$

which can be simplified using Taylor approximation for $\tilde{p}_i$ around $\tilde{p}_{i,0}$ to obtain an expression of toxic protein

$$\frac{\mathrm{d}\tilde{p}_i}{\mathrm{d}t} = -\rho \sum_{j=1}^{N} L_{ij} (\tilde{p}_i - \tilde{p}_{i,0}) + \left[\frac{k_{12} k_0}{k_1} - \tilde{k}_1 \right] (\tilde{p}_i - \tilde{p}_{i,0}) - \frac{k_{12}^2 k_0}{k_1^2} (\tilde{p}_i - \tilde{p}_{i,0})^2. \tag{5.6}$$

We have a Fisher-KPP expression of toxic protein dynamics

$$\frac{\mathrm{d}\tilde{p}_i}{\mathrm{d}t} = -\rho \sum_{j=1}^{N} L_{ij} (\tilde{p}_j - \tilde{p}_{j,0}) + \alpha (\tilde{p}_i - \tilde{p}_{i,0}) \left(1 - \frac{\tilde{p}_i - \tilde{p}_{i,0}}{\mathcal{K}} \right), \quad i = 1, ..., N, \tag{5.7}$$

where

$$\mathcal{K} = \frac{\alpha k_1^2}{k_0 k_{12}^2}, \quad \alpha = \frac{k_0 k_{12}}{k_1} - \tilde{k}_1. \tag{5.8}$$

We suppose $\alpha = \alpha(\tilde{k}_1)$ and $\mathcal{K} = \mathcal{K}(\tilde{k}_1)$. Further, we assume that the clearance of toxic proteins at node i, $\tilde{k}_{1,i}$, can vary in different brain regions due to toxic protein accumulation implying unequal regional growth rates $\alpha^{(i)} = \alpha(\tilde{k}_{1,i}) \neq \alpha(\tilde{k}_{1,j}) = \alpha^{(j)}$ and carrying capacities $\mathcal{K}_i = \mathcal{K}(\tilde{k}_{1,i}) \neq \mathcal{K}(\tilde{k}_{1,j}) = \mathcal{K}_j$ for $i \neq j$. With this assumption, we define $\tilde{c}_i = (\tilde{p}_i - \tilde{p}_{i,0})$ and write our model

$$\frac{\mathrm{d}\tilde{c}_i}{\mathrm{d}t} = -\rho \sum_{j=1}^{N} L_{ij}\tilde{c}_j + \tilde{\alpha}\tilde{c}_i\left(\mathcal{K}_i - \tilde{c}_i\right), \quad i = 1, ..., N, \tag{5.9}$$

where $\tilde{\alpha} = \alpha^{(i)}/\mathcal{K}_i = k_0 k_{12}/k_1$ is a global constant growth rate.

5.2.2 Heterogeneous growth rates

Another way of introducing heterogeneity is by varying the regional growth parameters. We suppose that toxic protein clearance can vary in different brain regions due to protein accumulation implying different local protein growth rates. We define $c_i = (\tilde{p}_i - \tilde{p}_{i,0})/\mathcal{K}$ and write equation (5.7) into

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} = -\rho \sum_{j=1}^{N} L_{ij}c_j + \alpha_i c_i(1 - c_i), \quad i = 1, ..., N. \tag{5.10}$$

where α_i denotes a local growth rate for node i. We also use the standard graph Laplacian $\mathbf{L}$ given by (5.4) for toxic protein transport on the brain network.

5.2.3 Dynamics of heterogeneous systems

We have introduced mathematical models that include heterogeneity. In this section, we want to show the effects of heterogeneous parameters in the system. We use a 5-node undirected graph depicted in Figure 5.1 where its adjacency graph is given by

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1/8 & 0 & 0 \\ 1 & 0 & 1/2 & 0 & 0 \\ 1/8 & 1/2 & 0 & 1/2 & 1/8 \\ 0 & 0 & 1/2 & 0 & 1/1000 \\ 0 & 0 & 1/8 & 1/1000 & 0 \end{bmatrix}. \tag{5.11}$$

We use uniformly random carrying capacities and growth rates between 0.8 and 1.2 where $\mathcal{K}_1 = 0.968, \mathcal{K}_2 = 1.189, \mathcal{K}_3 = 0.842, \mathcal{K}_4 = 0.918, \mathcal{K}_5 = 0.909$ for (5.9) and $\alpha_1 = 0.968, \alpha_2 = 1.189, \alpha_3 = 0.842, \alpha_4 = 0.918, \alpha_5 = 0.909$ for (5.10). We choose node 4 as the seeding node for both systems where $\tilde{c}_4(0) = 0.1$ for (5.9) and $c_4(0) = 0.1$ for (5.10) and zero otherwise.

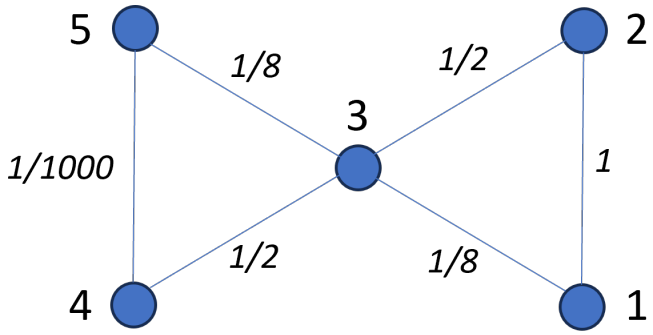


Figure 5.1: A 5-node undirected network

Figure 5.2 shows the difference of protein propagation using Fisher-KPP models (5.9) and (5.10). For the simulation, we choose two different diffusion rates ρ. Figures on the left panel (5.2a,c) depict the concentration evolution when parameters are in the growth-dominated regime. The behaviour resembles the standard Fisher-KPP, but toxic protein concentrations at different nodes grow differently due to different growth rates. For example, node 2 (yellow). This node exhibits a quick growth towards the plateau compared to the other nodes. Propagation in Figure 5.2c shows a similar growth as in Figure 5.2a for all nodes. Figures on the left panel (5.2b,d) depict the concentration evolution when parameters are in *the diffusion-dominated regime* ($\rho/\alpha \gg 1$). Both plots show that the seed initially spreads to all nodes in the network and low-concentration toxic proteins grow together showing one sigmoidal curve. Despite different plateaus in equation (5.9), we still observe a similar behaviour, which spread first and then grow. We also observe that the systematic node-by-node propagation is broken when local growths or capacities are significantly different. For example, node 3 has $\alpha_5 = 3$ while node 5 has $\alpha_5 = 3$ (figure not shown).

5.3 Arrival times of heterogeneous systems

It has been argued that the heterogeneity increases the speed of the invasion process creating a smaller average arrival times [5]. Instead of the change in global metric, we are interested in the underlying mechanisms that affect the change in arrival times. We will explore further the arrival times of both systems in this section and derive a general framework that works for both cases.

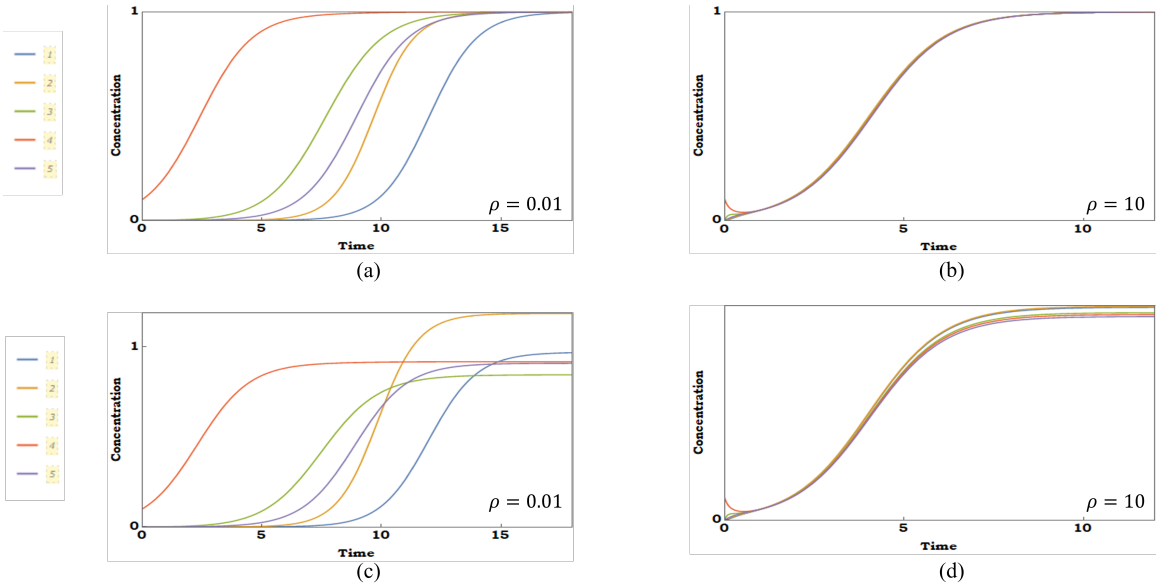


Figure 5.2: Fisher-KPP simulations on undirected setup in Figure 5.1. Top: results of equation (5.10) with $\tilde{\alpha} = 1$: (a) $\rho = 0.01$ and (b)$\rho = 10$. Bottom: results of equation (5.9): (c) $\rho = 0.01$ and (d) $\rho = 10$

5.3.1 Linear arrival times

The first approach to finding the arrival times is through a linearised system around the unstable healthy state, as discussed in Chapter 4. We follow the same process to obtain linear arrival time for heterogeneous systems. First, linearising (5.9) or (5.10) around the healthy state reduces the system to

$$\begin{aligned} \frac{\mathrm{d}c_i}{\mathrm{d}t} &= -\rho \sum_{j=1}^{N} L_{ij} c_j + \alpha_i c_i, \quad i = 1, ..., N, \\ c_i(0) &= q_i, \quad i = 1, ..., N, \end{aligned} \tag{5.12}$$

where $\alpha_i = \tilde{\alpha}\mathcal{K}_i$ for (5.9). For simplicity, we drop the tilde notation from (5.9). At this level, systems (5.9) and (5.10) share a similar form of linear progression.

Let $\mathbf{c}(t) = (c_1, ..., c_2)$, $\mathbf{q} = \mathbf{c}(0) = (q_1, ..., q_N)$, and $\mathbf{M} = -\rho\mathbf{L} + \mathcal{A}$, where $\mathcal{A}$ denotes the diagonal matrix of growth rates and $\mathbf{L}$ represents the graph Laplacian. In the case of equal growth rates, we have $\mathcal{A} = \alpha\mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix. For a linearised system, the solution is given as follows

$$\mathbf{c}(t) = e^{t\mathbf{M}}\mathbf{q}. \tag{5.13}$$

Since $\mathbf{M}$ is symmetric, there exists a set of eigenvectors $\mathbf{v}^{(i)} \in \mathbb{R}^N$ associated with eigenvalues $\Lambda_i \in \mathbb{R}$ for $i = 1, ..., N$ so that

$$\mathbf{M}\mathbf{v}^{(i)} = \Lambda_i \mathbf{v}^{(i)}, \quad i = 1, \ldots, N. \tag{5.14}$$

Then, the linear solution can be written explicitly as

$$\mathbf{c} = \sum_{i=1}^{N} \left(\mathbf{q} \cdot \mathbf{v}^{(i)} e^{t\Lambda_i} \right) \mathbf{v}^{(i)}, \tag{5.15}$$

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

$$\mu = \sum_{i=1}^{N} \left(\mathbf{q} \cdot \mathbf{v}^{(i)} e^{\widehat{\tau}_j \Lambda_i} \right) \mathbf{v}_j^{(i)}. \tag{5.16}$$

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

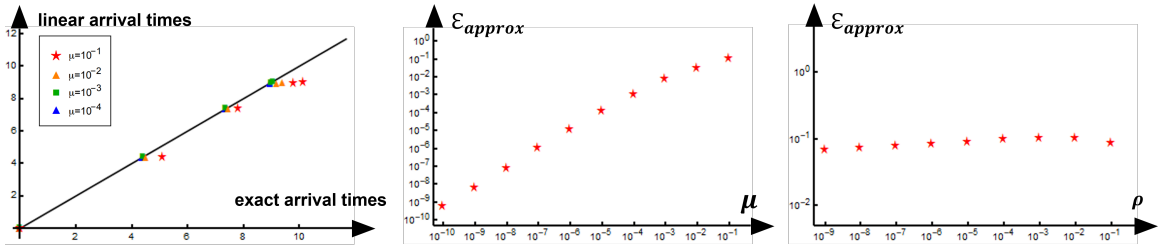
$$\mu/\beta = \sum_{i=1}^{N} \left(\mathbf{v}_j^{(i)} \mathbf{v}_s^{(i)} e^{\widehat{\tau}_j \Lambda_i} \right). \tag{5.17}$$

This approach is commonly used to compute arrival times, particularly in epidemiological problems. We note several features of the method. First, it is simple and provides a faithful approximation, although it requires a good initial guess. It is a robust and universal method that is valid for small initial values. Second, the linear estimates become a lower bound for the arrival times. The solutions of linear equations are always larger than the full system with nonlinear terms. Hence, the estimates are always smaller than the actual arrival times. Third, the linear times combine the sum of all possible paths resembling a diffusion operator. One can expand the matrix exponential as a Taylor series showing a superposition of powers of $\mathbf{L}$. The elements of the power matrix denote the sequence of edges taken from the seed node to the target node. Moreover, the linearisation method serves as an important foundation for other methods to derive arrival times.

We provide a comparison of the approximations with the numerical solutions for our 5-node network to assess their performance. We define the error based on the exact and approximate arrival times as follows

$$\mathcal{E}_{\text{approx}} = \frac{\sum_{i=1}^{N} |\widehat{\tau}_i - \tau_i|}{\sum_{i=1}^{N} \tau_i}. \tag{5.18}$$

Figure 5.3 shows the performance of linear arrival times using the network in Figure 5.1. We observe a few general features. First, linear times still become a quite good estimate of the arrival times. The solution combines different growths and diffusion processes using the multiple paths on the network. This feature allows the exponential growth to capture the real early dynamics of systems (5.9) and (5.10).



(a) (b) (c)

Figure 5.3: Simulations on heterogeneous system using equation (5.12) with $\alpha_1 = 0.968, \alpha_2 = 1.189, \alpha_3 = 0.842, \alpha_4 = 0.918, \alpha_5 = 0.909$. (a) linear time vs arrival time for $\rho = 0.01$, (b) error of linear times with varying $\mu = \beta$ for $\rho = 0.01$, and (c) error of linear times with varying ρ for $\mu = \beta = 0.1$.

As $\mu \to 0^+$, the approximation converges to the exact value and the nonlinear terms are negligible. Second, linear times become the lower bound of the arrival times. However, this approach's accuracy decreases when nodes have small growth rates as the linear solution grows way too fast compared to the original system. Third, we need a good initial guess to solve the transcendental equation. We can utilise the Lambert method to provide initial guesses although this method may fall apart in estimating the arrival times when small growth rates occur in the system.

5.3.2 Lambert times

We define a notion of time between nodes for this problem. It is the duration for an initial seed to propagate to a neighbouring node while ignoring other nodes in the network. We 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 (5.16) 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 $\mathbf{L}$, i.e. $\lambda_1 = 0$ and $\lambda_2 = 2A_{ij}$, and (5.17) with $\mu = \beta$ yields an equation for the arrival time at node j, from an initial seeding at node i (or vice-versa) and expanding for small ρ, we obtain

$$\frac{\rho A_{ij} 2 e^{\frac{t}{2}(\alpha_i+\alpha_j)} \sinh \frac{t}{2}\sqrt{(\alpha_i-\alpha_j)^2}}{\sqrt{(\alpha_i-\alpha_j)^2}} = 1. \tag{5.19}$$

Using Taylor expansion of order 1 for the sinh term gives

$$\rho A_{ij} t e^{\frac{t}{2}(\alpha_i+\alpha_j)} = 1. \tag{5.20}$$

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 $t = W_0(z)$):

$$t_{ij} = \frac{2}{\alpha_i + \alpha_j} W_0\left(\frac{\alpha_i + \alpha_j}{2\rho A_{ij}}\right), \qquad i \neq j. \tag{5.21}$$

We note that since the network is undirected, we have $A_{ij} = A_{ji}$ which implies that $t_{ij} = t_{ji}$ and we obtain a limiting form of equation (4.18) when $\alpha_i = \alpha_j$. We refer to t_{ij} as the *Lambert edge distance*. From this pairwise magnitude between neighbouring 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

$$W_{ij} = \begin{cases} \displaystyle\sum_{k=0}^{n-1} t_{\gamma_k,\gamma_{k+1}} = \sum_{k=0}^{n-1} \frac{2}{\alpha_{\gamma_k} + \alpha_{\gamma_{k+1}}} W_0\left(\frac{\alpha_{\gamma_k} + \alpha_{\gamma_{k+1}}}{2\rho A_{\gamma_k,\gamma_{k+1}}}\right), & i \neq j, \\ 0, & i = j, \end{cases} \tag{5.22}$$

which defines the *Lambert distance matrix* $\mathbf{W}$. This magnitude is a metric similar to the Lambert times in Chapter 4.

The same issue in estimating arrival times is that critical concentration may be different from the initial concentration, $\mu \neq \beta$, we also 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. In general, the time t_{ii} at which the solution to

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} = \alpha_i c_i \left(1 - \frac{c_i}{c_{i,\infty}}\right), \quad c_i(0) = \beta,$$

satisfies $c(t_{ii}) = \mu$ with equilibrium $c_{i,\infty}$. The solution to this problem is

$$t_{ii} = \frac{1}{\alpha_i} \log\left(\frac{(c_{i,\infty} - \beta)\mu}{(c_{i,\infty} - \mu)\beta}\right). \tag{5.23}$$

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

$$\tilde{\tau}_{ij} = t_{ii} + W_{ij}. \tag{5.24}$$

We can obtain simple expressions for the front solution by assuming that the dynamics after the arrival time is controlled by the local dynamics. This behaviour is strongly observed in Figure 5.2. We assume that, after this time, the effect of the graph Laplacian is neglected. More specifically, we assume that node i is chosen

as the seeding location and we can compute the equilibrium $\tilde{c}_{j,\infty}$, with initial seed $c_i(0) = \mu$, then the dynamics at node j is given by

$$\frac{\mathrm{d}\tilde{c}_j}{\mathrm{d}t} = \alpha_j \tilde{c}_j \left(1 - \frac{\tilde{c}_j}{\tilde{c}_{j,\infty}}\right), \quad \tilde{c}_j(\tilde{\tau}_{ij}) = \mu, \tag{5.25}$$

where $\tilde{\tau}_{ij}$ is the Lambert arrival time (4.22). The *Lambert solution*

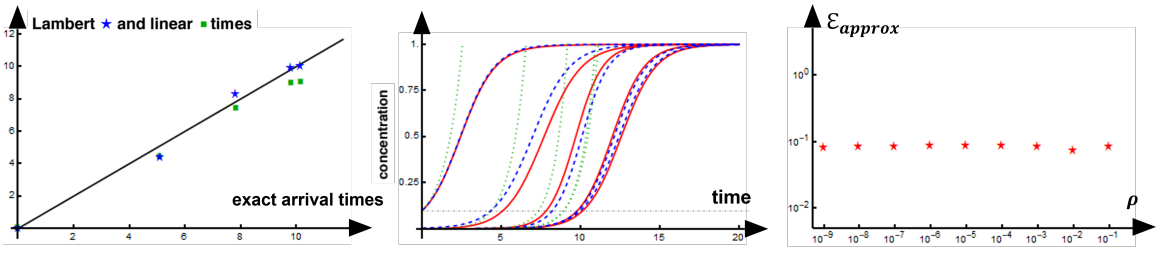
$$\tilde{c}_j = \frac{\tilde{c}_{j,\infty}\mu e^{\alpha_j t}}{\mu e^{\alpha_j t} - (\mu - \tilde{c}_{j,\infty})e^{\alpha_j \tilde{\tau}_{ij}}}, \quad j = 1, \ldots, N, \tag{5.26}$$

is therefore an approximation of (5.9) or (5.10).

We provide the Lambert edge distance (5.21), Lambert length (5.22), and Lambert solution (5.26) for the undirected graphs in Figure 5.1. We use $\alpha_1 = 0.968, \alpha_2 = 1.189, \alpha_3 = 0.842, \alpha_4 = 0.918, \alpha_5 = 0.909$, and $\rho = 0.01$. The Lambert edge distance and Lambert distance matrices for the graph are given in (5.27)

$$[t_{ij}] = \begin{bmatrix} \infty & 3.19 & 5.5 & \infty & \infty \\ 3.19 & \infty & 3.88 & \infty & \infty \\ 5.5 & 3.88 & \infty & 4.35 & 5.66 \\ \infty & \infty & 4.35 & \infty & 10.1 \\ \infty & \infty & 5.66 & 10.1 & \infty \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} 0. & 3.19 & 5.5 & 9.85 & 11.2 \\ 3.19 & 0. & 3.88 & 8.23 & 9.54 \\ 5.5 & 3.88 & 0. & 4.35 & 5.66 \\ 9.85 & 8.23 & 4.35 & 0. & 10. \\ 11.2 & 9.54 & 5.66 & 10. & 0. \end{bmatrix}, \tag{5.27}$$

where we use the symbol ∞ to characterise pairs of nodes that are not connected by an edge (implying infinite Lambert edge length). Figure 5.4 demonstrates the corresponding arrival times and Lambert solution. The Lambert method gives an improvement to the arrival time and solution. The approach is robust as the transport rate ρ varies as depicted in Figure 5.4c.



(a) (b) (c)

Figure 5.4: Simulations on heterogeneous system using equation (5.17) with $\alpha_1 = 0.968, \alpha_2 = 1.189, \alpha_3 = 0.842, \alpha_4 = 0.918, \alpha_5 = 0.909$, and $\rho = 0.01$. (a) Lambert time vs arrival time, (b) exact (red), Lambert (dashed blue), and linear (dotted green) solutions, and (c) error of Lambert times with varying ρ for $\mu = \beta = 0.1$.

We note several features of the method. First, the Lambert distance is an intrinsic property of heterogeneous systems. Despite different local features between nodes,

the distance between two distinct nodes satisfies metric. Second, the shortest path distance based on the Lambert edge distance dominates the propagation when $|\alpha_i - \alpha_j|$ is small. However, the ability to estimate arrival times breaks down when the difference is large enough. Third, it is still a good first guess to the linear times.

5.3.3 Nonlinear asymptotic times

We previously assumed that the propagation on the growth-dominated regime follows the node-by-node process according to the Lambert times. The initial nonlinear asymptotic method as in Chapter 4 strongly takes into account the process of the Lambert method. However, this process breaks down when the system has small growth rates. It results in a different sequence of propagation. We assume that the dynamics of a node can be found by a direct connection between the seed and that node. We consider the canonical case and the *ansatz* is given by

$$\tilde{c}_i = \frac{\kappa_i K_i e^{\alpha_i t}}{\kappa_i + K_i \left(e^{\alpha_i t} - 1\right)}, \quad i = 1, \ldots, N. \tag{5.28}$$

where κ_i is the new diseased state at node i due to local capacities $\mathcal{K}_i$ for $i = 1, ..., N$ in (5.9). We use again a nonlinear version of the method of variation of constants as introduced in Chapter 4.

$$K_i = A_i + \rho B_i(t). \tag{5.29}$$

With this ansatz, we can solve for the unknown function B_i for all nodes by integrating once from the seeding node. To lowest order $\mathcal{O}(\rho^0)$, the solution at the seeding node with initial condition $c_s(0) = \beta$ is simply given by the original unperturbed solution

$$\tilde{c}_s = \frac{\kappa_s \beta e^{\alpha t}}{\kappa_s + \beta \left(e^{\alpha t} - 1\right)}, \tag{5.30}$$

and $A_i = 0$ for all $i \neq s$. We then need to solve

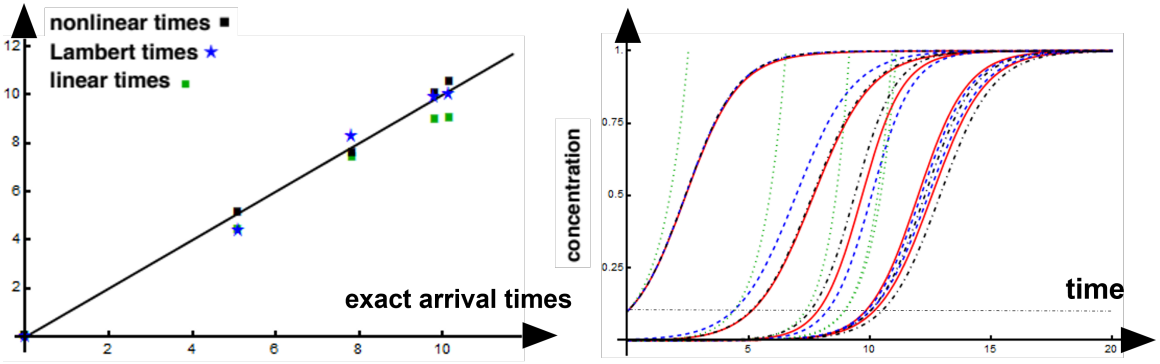
$$\frac{\mathrm{d}B_i}{\mathrm{d}t} = \sum_{s=1}^{\mathcal{S}} \frac{M_{is}}{\kappa_i} e^{-\alpha_i t} \tilde{c}_s, \qquad B_i(0) = 0, \qquad i \neq s, \tag{5.31}$$

where M_{is} must be determined. As we have mentioned, the dynamics of heterogeneous systems may break the node-by-node propagation principle following the shortest path. We replace the long connectivity between seed and node with a 'virtual' direct short edge between them. We note that, in general, the linear times capture the slow dynamics due to small growth rates and naturally combine multiple paths on the

network. Specifically, the canonical linear times $\tilde{t}_{is}$ summarises the canonical linear dynamics of (5.9) or (5.10). We also note that (5.21) captures a similar behaviour of linear dynamics between seed and node. We then define the replacement edge M_{is}

$$M_{is} = \frac{e^{-\frac{\tilde{t}_{is}}{2}(\alpha_i+\alpha_s)}}{\rho \tilde{t}_{is}}. \tag{5.32}$$

The solution of (5.31) are given either by a Hypergeometric function ${}_2F_1(a, b; c; z)$ when $\alpha_i \neq \alpha_s$ or by a natural logarithm function, when $\alpha_i = \alpha_s$.



(a) (b)

Figure 5.5: Simulations on heterogeneous system using nonlinear method with $\alpha_1 = 0.968, \alpha_2 = 1.189, \alpha_3 = 0.842, \alpha_4 = 0.918, \alpha_5 = 0.909$, and $\rho = 0.01$. (a) Nonlinear time vs arrival time for $\rho = 0.01$, (b) exact (red), nonlinear (dash-dotted black), Lambert (dashed blue), and linear (dotted green) solutions.

Figure 5.5 shows the arrival times and solution of the Fisher-KPP (5.10) on the undirected network in Figure 5.1. The *nonlinear asymptotic solutions* satisfy the initial values and the asymptotic behaviour for a long time. We compute the errors of arrival times in Figure 5.5 and the errors decrease as we move from the linear times, Lambert times, to nonlinear times: 1.01×10^{-1}, 4.1×10^{-2}, and 2.6×10^{-2}. Notably, the nonlinear method significantly improves the Lambert times and Lambert solutions.

We note several features of the nonlinear asymptotic approach for varying kinetics. First, the solution accommodates both heterogeneous growth and balancing between kinetics and transport. It satisfies both initial conditions and asymptotic behaviour. Second, different local capacities or growths can break down the step-by-step propagation. Instead, we use a replacement connectivity that connect directly two distinct nodes. The new 'virtual' weight can be obtained by the linear times. Third, the solution is either a logarithmic or hypergeometric function according to the growths

between the seed and target nodes. The solution exhibits a simple expression compared to the step-by-step process introduced in Chapter 4.

5.4 Application to neurodegenerative disease

In this section, we apply methods to estimate the arrival time to the study of the propagation of toxic protein transported on a structural connectome and explore our models with respect to staging. We assume that the heterogeneity comes from the degree of nodes and local capacity data. We use the scale-33 structural connectome that has 83 nodes. Each node represents a brain region. To produce staging, we use the six Braak stages where the corresponding regions of each stage are based on the Braak staging given by ADNI.

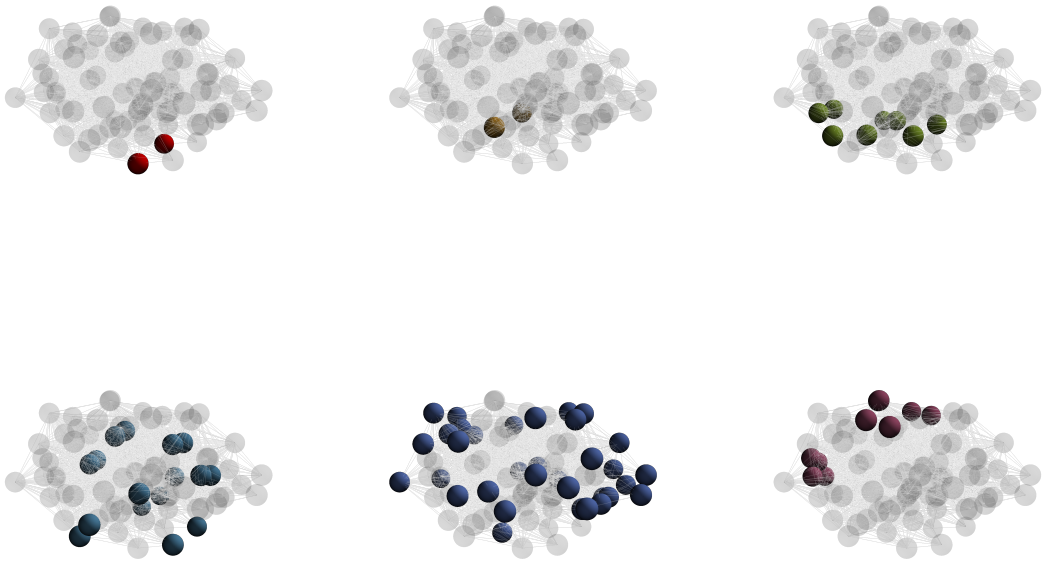


Figure 5.6: Braak regions of interest according to ADNI. Top: Braak I (left), Braak II (middle), Braak III (right). Bottom: Braak IV (left), Braak V (middle), Braak VI (right).

Brain regions corresponding to Braak stages are depicted in Figure 5.6. The six stages are a sequence of τP hierarchical propagation in the brain [17, 18]. The starting location is the entorhinal cortex associated with Braak stage I. The expected

progressive pattern of dynamics follows the Braak staging. For the scale-33 brain connectome, the entorhinal cortex corresponds to nodes 27 and 68.

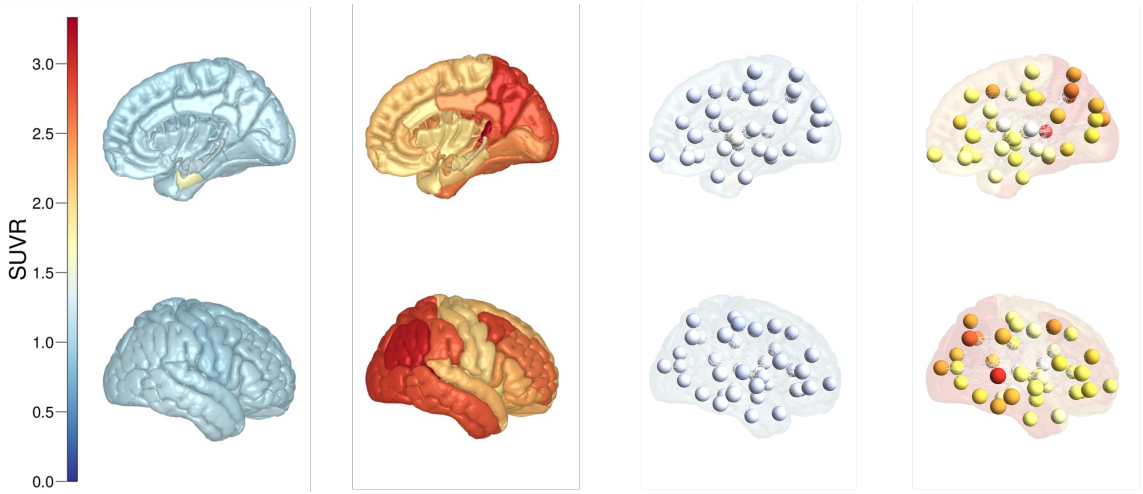


Figure 5.7: Local τP SUVR scores across the brain. Brains on the first two columns (left) respectively show the minimum and maximum level of τP SUVR scores. The last two columns (right) show their associated networks.

PET scans of τP suggest the existence of lower and upper bound of toxic τP concentrations. From the data, we infer the local parameters for each node representing regional baseline values for a healthy state and carrying capacities for a late-stage AD state. The parameters used in this study include information about the PET measurement process and AD, such as off-target binding and regional variations in production dynamics. These parameters allow us to model τP SUVR score directly. The Gaussian mixture modelling approach developed in [180] is used to obtain these parameters. To estimate the fixed parameters for $c_{i,0}$ and $c_{i,1}$ representing the minimum and maximum τP SUVR scores, respectively, we fit a two-component Gaussian mixture model to population-level data of regional SUVR. For regions where a reliable measure of τP SUVR can be obtained, we expect to see two separate distributions, a τP$-$ distribution capturing the expected τP load in a given region, and a τP+ distribution describing the pathological τP load [180]. Using the fitted Gaussian mixture models, we approximate $c_{i,0}$ as the mean of the τP$-$ distributions and $c_{i,1}$ as the 99th percentile of the τP+ distributions. Due to off-target binding, it is not possible to obtain τP PET signal for subcortical regions [32, 158]. Therefore, these regions are excluded from our model, leaving a total of 72 regions. The multi-tau cohort of AV1451 PET data detailed in [180] is used for ADNI. Figure 5.7 shows the regional minimum and maximum capacities in the brain. If we define the regional capacity

difference, the caudate has the smallest values while the middle temporal has the largest values.

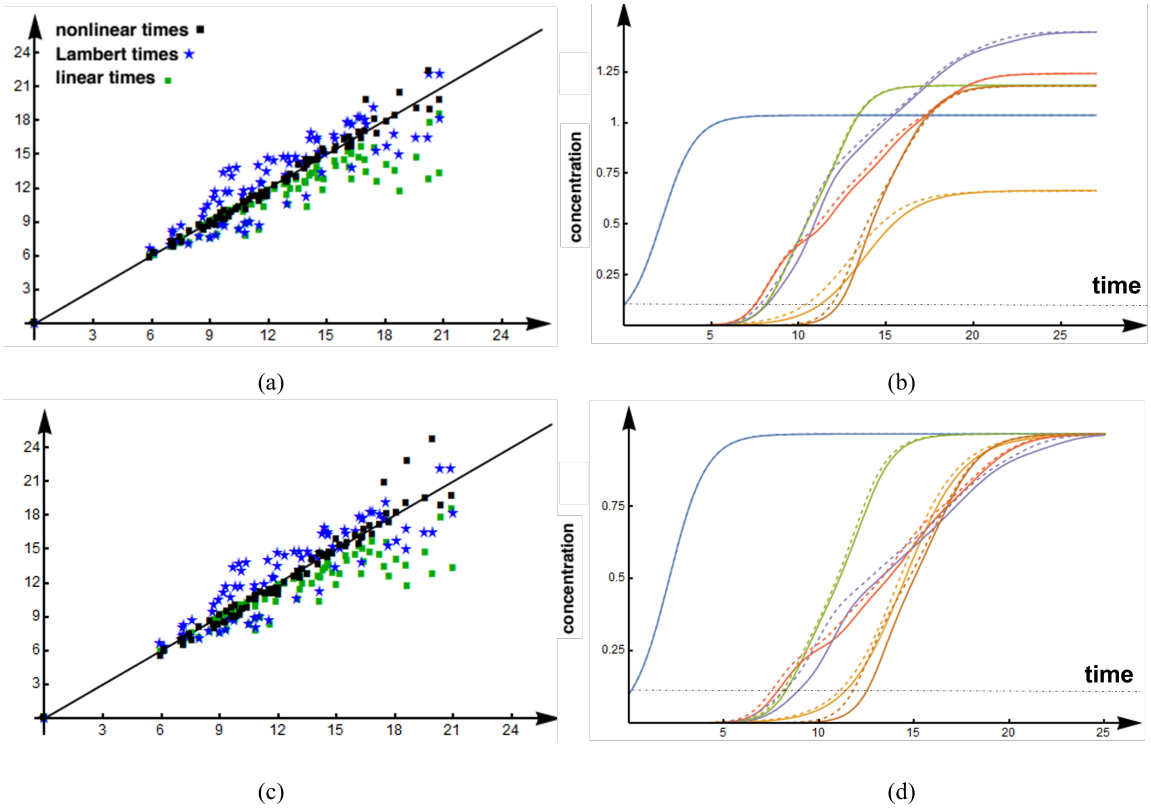
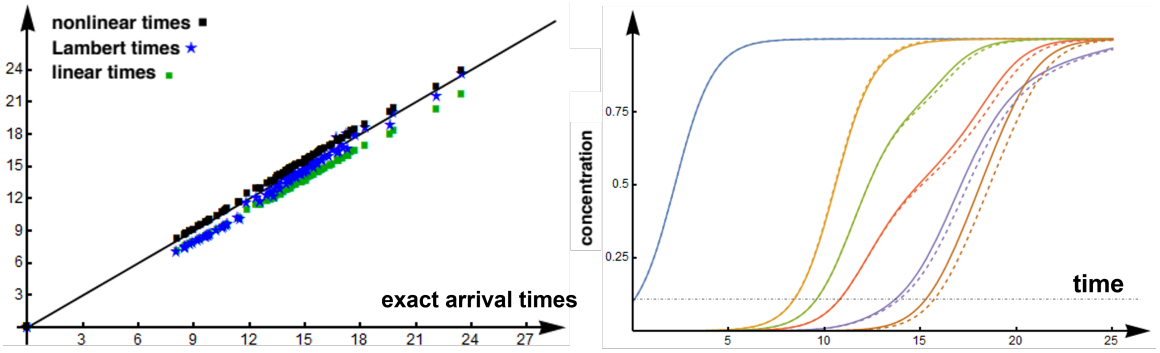


Figure 5.8: Simulations on varying kinetics using nonlinear method. Top: simulation for (5.9). Bottom: simulation for (5.10). (a,c) Nonlinear times vs arrival times for $\alpha = 1$ and $\rho = 0.01$, (b,d) exact (bold), nonlinear (dashed).

5.4.1 Tau protein propagation

Figure 5.8 demonstrates the application of our methods to (5.9) and (5.10). Since we are interested in similar arrival times between two equations, we use $\alpha_i = \alpha\mathcal{K}_i$ where $\mathcal{K}_i = c_{i,1} - c_{i,0}$. The errors decrease as we move from the linear times, Lambert times, to nonlinear times: 1.3×10^{-1}, 1.2×10^{-1}, and 6.19×10^{-2} for (5.9) and 1.35×10^{-1}, 1.16×10^{-1}, and 7.56×10^{-2} for (5.10). The nonlinear solution provides a good approximation although the asymptotic values vary. This approach also works for the propagation on the undirected network as shown by Figure 5.9 where the errors decrease as we move from the linear times, Lambert times, to nonlinear times: 9.89×10^{-2}, 4.98×10^{-2}, and 2.05×10^{-2}.

Our results show that the nonlinear asymptotic method provides a good, and best overall, estimate for both arrival times and dynamics. In particular, the approach per-



(a) (b)

Figure 5.9: Simulations on homogeneous using the nonlinear method. (a) Nonlinear time vs arrival time for $\alpha = 1$ and $\rho = 0.01$, (b) exact (bold), nonlinear (dashed).

fectly captures the average dynamics. The method, however, needs a crucial input that depends on the numerical simulation due to the emergence of new asymptotic values. The assumption that works on the heterogeneous system can also be implemented for the propagation on undirected networks.

5.4.2 Staging of tau protein

We have applied each approximation method to heterogeneous systems. We are interested in exploring the influence of heterogeneity in the system towards the resulting patterns. We assess each model systematically through staging analysis in the growth-dominated regime ($\rho/\alpha \ll 1$) using the braid diagram and braid surface as introduced in Chapter 3.

To understand the influence of parameter changes to staging, we perturb local capacities of (5.9) and local growths of (5.10). Therefore, we write (5.9) as

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} = -\rho \sum_{j=1}^{N} L_{ij} c_j + \alpha c_i (1 + \varepsilon \mathcal{K}_i - c_i), \quad i = 1, ..., N, \tag{5.33}$$

where $\varepsilon \mathcal{K}_i \ll 1$ for all i. Next, we write (5.10) as

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} = -\rho \sum_{j=1}^{N} L_{ij} c_j + (\alpha + \varepsilon \alpha_i) c_i (1 - c_i), \quad i = 1, ..., N, \tag{5.34}$$

where $\varepsilon \alpha_i \ll 1$ for all i. For both (5.33) and (5.34), α is a global growth rate while ε is a tuning parameter. We use initial values $c_s = \beta$ for seeding nodes s and zero otherwise.

We note that the asymptotic behaviour over the long time of heterogeneous systems can be different from the standard dynamics in which the asymptotic behaviour is the Kernel of the graph Laplacian. We introduce a new magnitude in the range of 0 and 1 out of J regions $\Omega_1, ..., \Omega_J$. In each of these regions, we define

$$C_j(t) = \frac{1}{\sigma_{\Omega_j}} \sum_{i \in \mathcal{V}_j} c_i(t), \quad j = 1, ..., J, \tag{5.35}$$

where

$$\sigma_{\Omega_j} = \sum_{i \in \mathcal{V}_j} c_{i,\infty},$$

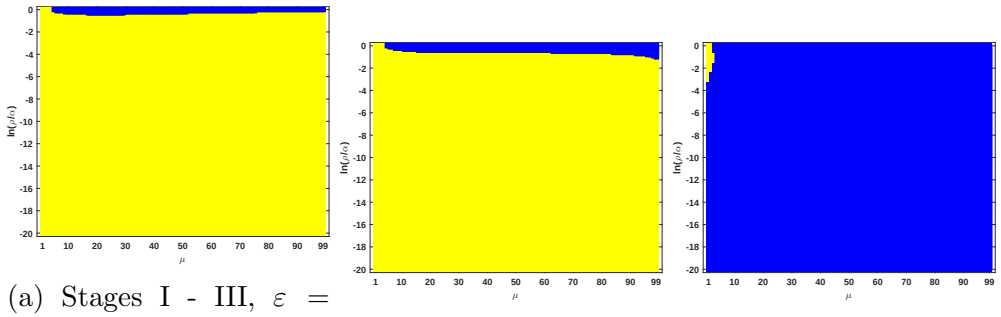
where $\mathcal{V}_j$ is the collection of nodes defining region Ω_j with N_j nodes and $c_{i,\infty}$ represents the nontrivial asymptotic concentration at node i. The average value (5.35) accommodates the new asymptotic without forcing each node to evolve between 0 and 1. In our simulations, the first region Ω_1 is the bilateral entorhinal cortex with initial conditions $c_i(0) = 0.005/N_1$ for $i \in V_1$ and zero otherwise. We determine two staging sequences of the first three stages and the last three stages. Using this staging, we can see clearly how heterogeneity alters the patterns resulting from a homogeneous system. For a pair of parameter ρ/α and threshold value, we assign a colour label. The assigned colours are shown in Table 5.1. The expected canonical Braak staging of τP is given by progressive sequences I $\rightarrow$ II $\rightarrow$ III for the first three Braak stages and IV $\rightarrow$ V $\rightarrow$ VI for the last three stages. The yellow colour indicates the canonical staging sequence. We summarise all computed colours in our simulations into braid surfaces.

All observed computational staging patterns				
Braak regions I, II, and III		I $\rightarrow$ II $\rightarrow$ III *		I $\rightarrow$ III $\rightarrow$ II
		Beginning with stage II		Beginning with stage III
Braak regions IV, V, and VI		IV $\rightarrow$ V $\rightarrow$ VI*		IV $\rightarrow$ VI $\rightarrow$ V
		V $\rightarrow$ IV $\rightarrow$ VI		V $\rightarrow$ VI $\rightarrow$ IV
		VI $\rightarrow$ IV $\rightarrow$ V		VI $\rightarrow$ V $\rightarrow$ IV

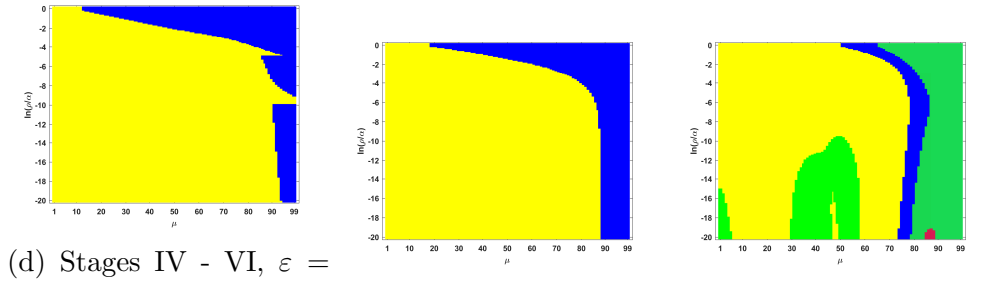
Table 5.1: Color coding for braid surfaces of τP propagation. * Progressive computational Braak staging.

We compute staging when heterogeneity occurs from the brain regional capacities. In this case, we simulate (5.33) and (5.34) on undirected brain connectome with the diffusive weights by tuning the local parameter gradually. We are interested in systems that have similar arrival times. Therefore, we determine that $\alpha_i = \alpha \mathcal{K}_i$ in

(5.34). The values of $\mathcal{K}_i$ are the regional capacity differences between the maximum and minimum levels of τP SUVR scores. Without loss of generality, we choose α to be the order of $\mathcal{O}(1)$. We use $\varepsilon = -0.1$ as the lowest perturbation value to ensure that the local capacity or growth has a positive value while the largest is $\varepsilon = 1$. Figure 5.10 shows staging patterns from (5.33). Figure 5.11 shows staging patterns from (5.34). In general, we observe that staging is consistent as the local capacity or local growth rates are reduced proportionally to the local capacity difference. However, staging tends to alter as local capacity or local growth increases. The most apparent alterations occur for the first three stages, while the last three-stage staging is strongly influenced at higher threshold values. Interestingly, the concentration in region III tends to accumulate faster than in region II and staging is stable for the whole evolution. Further results with various ε values are shown in the Appendix 7.2.2. Our results suggest that the local capacity or local growth affects staging in the first three stages. Although local heterogeneity remains unclear with the possibility of regional vulnerability or the breakdown of regional clearance, higher heterogeneous rates indeed change the pattern of τP propagation across the brain.

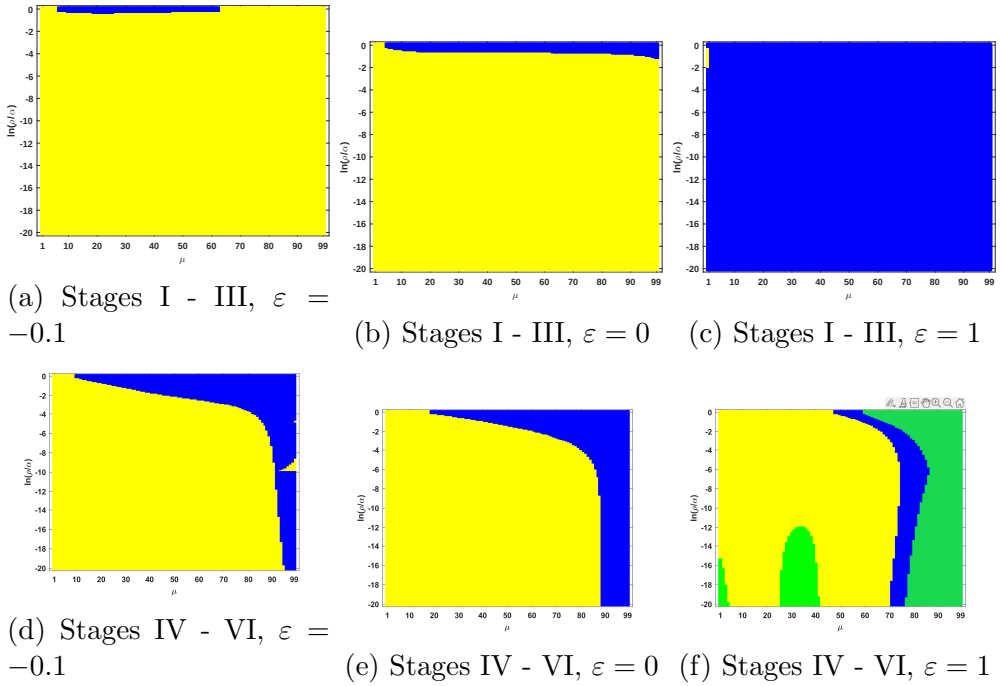


(a) Stages I - III, $\varepsilon = -0.1$ (b) Stages I - III, $\varepsilon = 0$ (c) Stages I - III, $\varepsilon = 1$



(d) Stages IV - VI, $\varepsilon = -0.1$ (e) Stages IV - VI, $\varepsilon = 0$ (f) Stages IV - VI, $\varepsilon = 1$

Figure 5.10: Observed computational (deterministic) connectome τP seed staging; Stages I - III (top) and stages IV - VI (bottom). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (5.33).



(a) Stages I - III, $\varepsilon = -0.1$ (b) Stages I - III, $\varepsilon = 0$ (c) Stages I - III, $\varepsilon = 1$

(d) Stages IV - VI, $\varepsilon = -0.1$ (e) Stages IV - VI, $\varepsilon = 0$ (f) Stages IV - VI, $\varepsilon = 1$

Figure 5.11: Observed computational (deterministic) connectome τP seed staging; Stages I - III (top) and stages IV - VI (bottom). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (5.34).

Our results show that heterogeneity in the system influence the landscape of observable staging patterns. In particular, pattern alterations occur when parameter values change considerably. The most apparent is the influence of large heterogeneous values.

5.5 Concluding remark

In this chapter, we have presented heterogeneous systems that may arise in the study of τP propagation on the brain connectome. Our objective is to explore any possible models that emerge from the fact of nontrivial toxic τP concentration in healthy patients. As a follow-up to Chapters 3 and 4, we study heterogeneity and its influence on the overall dynamics and estimates.

Our models on τP propagation emphasise the role of growth and transport. In this case, heterogeneity is included in the kinetics, which contains local parameters. Data suggest that healthy individuals can build up different τP accumulation in different brain regions. This finding implies a shift to the toxic protein concentration baseline. Further assumptions of the involvement of regional vulnerability or heterogeneous

clearance lead to varying kinetics given by different local capacities or local growth rates. The emergence of heterogeneous features alters the asymptotic behaviour and the Kernel of the graph Laplacian is not an equilibrium. Staging analysis suggests that heterogeneity affects the resulting staging patterns, particularly when heterogeneous growth rates are large. The study of arrival times suggests small local capacities or growths can delay the arrival times far from the homogeneous case. We derive three estimates for the arrival times to understand the underlying mechanism of propagation.

The first approach is to linearise the full system around the healthy state. We then obtain the solution of protein concentration and solve the transcendental equation to obtain linear arrival times. The method is universal for varying kinetics. However, the estimate is frail for nodes having small local capacities or growth rates.

The second approach is to derive an estimate between two adjacent nodes and to take the sum over a path. The path is likely the shortest path since the sum of estimates gives the smallest value and dominates the propagation. The latter feature breaks down for varying kinetics, especially when the parameter is small or large enough. This condition allows to connect directly seeding nodes to target nodes. Despite, its limitation, the Lambert method is computationally cheap and independent of initial values.

The last method is an improvement on Lambert solution by including nonlinearity in the solution. The approach takes into account a small parameter similar to the approach in Chapter 4. The emergence of new asymptotic values is also considered in the solution. We address the problem found in the Lambert method for heterogeneous capacity or growth by introducing a 'virtual' direct edge between seeding nodes and target nodes whose weights are defined from the linear arrival times. Therefore, the solution can be in a logarithm or hypergeometric function. However, we still need information on asymptotic values from the numerical computation of the full system. This method shows a promising estimate despite its complexity and long expression.

Our study on staging and arrival time problems allows us to explore the influence of heterogeneity on the propagation process from a seed to its neighbouring nodes. To some extent, heterogeneity influences the resulting staging sequences. The dominating factor of invasion depends on the shortest path as shown by the Lambert method and nonlinear asymptotic approach. However, this characteristic is different in the heterogeneous kinetics. Our exploration of methods for estimating arrival times shows that these methods can accommodate heterogeneous Fisher-Kolmogorov-Petrovsky-Piskunov dynamical processes. In all, staging and arrival time problems

are essential in understanding the complex coupled process from both computational and analytical points of view.

Chapter 6

Directed Network System and Application to Neurodegenerative Disease

6.1 Overview

6.1.1 Directional bias in neurodegeneration

The structure of the central nervous system that controls memory, cognition, bodily regulations and functions is complex. Multiple studies have suggested that anatomical differences in the brain underlie behavioural differences between genders [83, 103, 166]. For example, males have a higher percentage of white matter containing myelinated axonal fibres and cerebrospinal fluid. A study using the brain connectome has revealed that average females have more neuronal edges showing better expander graphs between hemispheres [166]. This difference is believed to be influenced by the development of brain connectivity during different stages [83]. In addition, the connectivity is directed by the signal propagation from the neuronal soma to the end of the axon [112]. This process creates the directionality of cerebral connections.

Despite the brain's complexity, the directionality in the brain has been studied and translated into directed connectomes. This advantage allows scientists to analyse the fundamental organisation and integration of brain networks as a whole rather than investigating individual regions or isolated tracts [173]. A recent study incorporating macaque's directed-connectome has suggested that their study provides a theoretical framework of large-scale brain mechanisms in which self-sustained brain memory activities may emerge when multiple region interaction is considered [118]. In the case of neurodegenerative disease, studies have suggested that the transport of toxic proteins is influenced by neuronal directional bias [109, 174]. A seminal study has

suggested that the directionality in the brain plays an important aspect in transmission and increases the plausibility of data when compared to the model [131]. These approaches that involve directed networks are essentially predictive and are built in a data-driven fashion without probing more into the influence of directions to transmission. The question is now to reveal the dynamics and related estimates rather than just applying the model to the available data. Here, we propose a study on directed networks to explore the effect of directionality on physical transport dynamics.

In network modelling of physical transport processes, the graph Laplacian is an essential approach for modelling linear coupling dynamics between nodes. There are two different perspectives for modelling the graph Laplacian on a symmetric adjacency configuration. The first perspective involves a random walk on a network where the nodes represent the states of stochastic processes, where a walker starting from a seed node may move to another neighbouring node with a certain probability [52]. This approach has many applications in neurodegeneration studies, where it can predict atrophy patterns in neurodegenerative diseases such as Alzheimer's and Parkinson's [145] and Supranuclear Palsy [131] through Laplacian eigenmodes or spreading dynamics. While this approach is appealing in data-driven modelling, it neglects some important properties in physical transport [144]. The second perspective involves a combinatorial approach that computes the difference between inflow and outflow at each node and naturally provides physical diffusive transport on an undirected topology (satisfying the overall mass conservation and Fick's property) [144]. The graph Laplacian has been used to model toxic protein propagation in brain networks [49, 172] and can capture notable dynamics of toxic protein accumulation when coupled with nonlinear kinetics under this configuration. [144, 172]. However, incorporating directed networks into the model may change the dynamics and it needs to be studied. The main question is, again, to understand the resulting dynamics when directed networks are considered. For instance, if two networks have slight connectivity differences, do they show significantly distinct dynamics? Despite the brain's directionality uncertainty, the question is to reveal the dynamics due to directed connections in a system and derive related estimates rather than just applying the model to available data. In this study, we propose exploring the effect of directionality on physical transport dynamics in a network.

6.1.2 Exploration of an invasion process on directed networks

In this study, we restrict our exploration to linear coupling dynamics between nodes and the modelling of neurodegenerative diseases [120]. Similar to Chapter 5, our study

here is to explore the dynamics and notable measures such as *the arrival time problem*, which measures the time it takes for a quantity of interest to reach a determined threshold at each node, and *the staging problem*, which determines the sequence of regional spreading throughout the brain at different thresholds and parameter values.

We consider a general spatiotemporal system with a form

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = \mathbf{f}(t, \mathbf{p}; \boldsymbol{\Theta}), \tag{6.1a}$$

$$\mathbf{p}(0) = \mathbf{q}, \tag{6.1b}$$

where $\mathbf{p}(t) = (p_1, ..., p_N)$ denotes a quantity, for example, a protein concentration, evolving on a connected and directed network $\mathcal{G}^{\mathrm{dir}} = (\mathcal{V}, \mathcal{E}^{\mathrm{dir}})$ with a collection of N nodes $\mathcal{V}$, representing regions of interest, and directed edge set $\mathcal{E}^{\mathrm{dir}}$, representing connections between these regions. The quantity $p_i(t)$ corresponds to the observed concentration in node v_i and $p_i(0) = q_i$ is the initial concentration at that node, $\mathbf{q} = (q_1, ..., q_N)$. The quantity $\boldsymbol{\Theta}$ represents the parameters of the system 6.1. We further assume that the dynamics of the system is such that starting from an initial condition, where all concentrations are taken to vanish except in a few nodes, and the system will evolve asymptotically to a state where all concentrations reach their maximal value.

We recall again the canonical arrival-time problem, in which 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 $\tau_i = \tau(\mu)$ such that $c_i(\tau_i) = \mu$ (with $\beta \leq \mu < 1$ so that $\tau_i \geq 0$ and finite with the equality $\tau_i = 0$ attained if only if $i = s$ and $\mu = \beta$).

Next, we recall the staging problem. Let Ω_j for $j = 1, ..., J$ be a nonoverlapping collection of nodes, such that $\Omega_j \subseteq \mathcal{V}$ and $\Omega_j \cap \Omega_k =$ if $j \neq k$. Let $T \in [0, 1]$ be an arbitrary threshold value. As the concentration evolves according to (6.1), we define the weighted average concentration P_j

$$P_j = \frac{1}{\sigma_{\Omega_j}} \sum_{i \in \Omega_j} p_i, \quad j = 1, ..., J,$$

where

$$\sigma_{\Omega_j} = \sum_{i \in \Omega_j} p_{i,\infty},$$

with $p_{i,\infty}$ represents the nontrivial asymptotic concentration at node i. We record the time when P_j first reaches the threshold T, $P_j(t_j) = T$. This process produces an ordering of the regions Ω_j according to the ordered arrival times called an observed staging pattern $\{\Omega_{j_1}, \Omega_{j_2}, ..., \Omega_{j_J}\}$. The generalised staging problem is to probe observable staging patterns subject to varying $\boldsymbol{\Theta}$.

6.2 Mathematical model using directed networks

We suppose a directed connected graph $\mathcal{G}^{\text{dir}} = \left(\mathcal{V}, \mathcal{E}^{\text{dir}}\right)$ with a collection of N nodes $\mathcal{V}$ representing regions of interest and edges $\mathcal{E}^{\text{dir}}$ representing directed connections between these regions. $\mathbf{A}^{\text{dir}} = [a_{ij}]$ represents the adjacency $N \times N$ matrix of a directed network $\mathcal{G}^{\text{dir}}$ and $\mathbf{L}^{\text{dir}}$ is the graph Laplacian on directed network $\mathcal{G}^{\text{dir}}$. We define the graph Laplacian as

$$\mathbf{L}^{\text{dir}} = \mathbf{D}^{\text{out}} - \left(\mathbf{A}^{\text{dir}}\right)^T, \tag{6.2}$$

where $\mathbf{D}^{\text{out}}$ is the outdegree matrix, a diagonal matrix with entries $D_{ii}^{\text{out}} = d_i^{\text{out}} = \sum_j a_{ij}$. We note some features from this graph Laplacian. First, if $\mathbf{1} \in \mathbb{R}^N$ is the vector one, then we find that the mass conservation is given by $\left(\mathbf{1} \cdot \mathbf{L}^{\text{dir}} = \mathbf{0}\right)$, but the graph Laplacian does not satisfy the Fick's property. Rather, we have a transport property $\mathbf{L}^{\text{dir}} \cdot \mathbf{v} = \mathbf{0}$, where $\mathbf{v}$ is the Kernel of $\mathbf{L}^{\text{dir}}$ and is not necessarily $\mathbf{v} = \mathbf{1}$. This property occurs due to asymmetric flow between, at least, two nodes. It means that the same nonzero quantity of mass at each node will not stop the transfer process between two nodes due to different reverse flows. Second, complex eigenvalues can emerge since the network is asymmetric. Suppose λ is the eigenvalues of $\mathbf{L}^{\text{dir}}$. Complex eigenvalues $\lambda \in \mathbb{C}$ can emerge from the defined graph Laplacian. We find that $\mathbf{L}^{\text{dir}}$ has eigenvalues with $\text{Re}(\lambda) \geq 0$ with one zero eigenvalue, in the case of connected graphs due to a nontrivial Kernel. The asymmetric connections affect the fundamental characteristics and spectra of the transport process given by the graph Laplacian on directed graphs. If $\mathbf{c}(t) = (c_1(t), ..., c_N(t)) \in \mathbb{R}^N$ represent protein concentrations at time t, then the protein transport dynamics is governed by

$$\begin{aligned} &\frac{\mathrm{d}\mathbf{c}}{\mathrm{d}t} = -\rho \mathbf{L}^{\text{dir}} \mathbf{c}, \\ &\mathbf{c}(0) = \mathbf{q}_0, \end{aligned} \tag{6.3}$$

with initial values $\mathbf{q} = (q_1, ..., q_N)$ and ρ denotes a mobility rate. The solution is given by

$$\mathbf{c}(t) = e^{-\rho t \mathbf{L}^{\text{dir}}} \mathbf{q}.$$

We take two directed networks that have different characteristics. We choose a weakly connected graph[5] and a strongly connected graph[6] depicted in Figure 6.1. The adjacency graph for the weakly connected graph is given by

[5] A weakly connected graph is a directed graph if one can reach any node from any other node by traversing edges in an undirected path [126].

[6] A strongly connected graph is a directed graph if it is feasible to reach any node from any other node by moving along edges in the direction(s) they indicate [126].

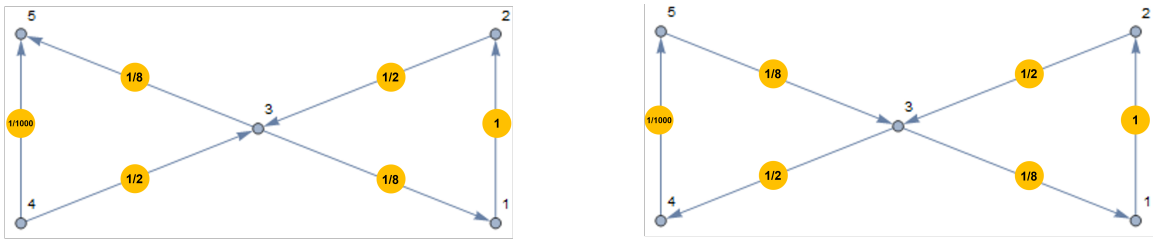


Figure 6.1: (a) Weakly connected graph and (b) Strongly connected graph.

$$\mathbf{A}^{\text{dir}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 1/8 & 0 & 0 & 0 & 1/8 \\ 0 & 0 & 1/2 & 0 & 1/1000 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{6.4}$$

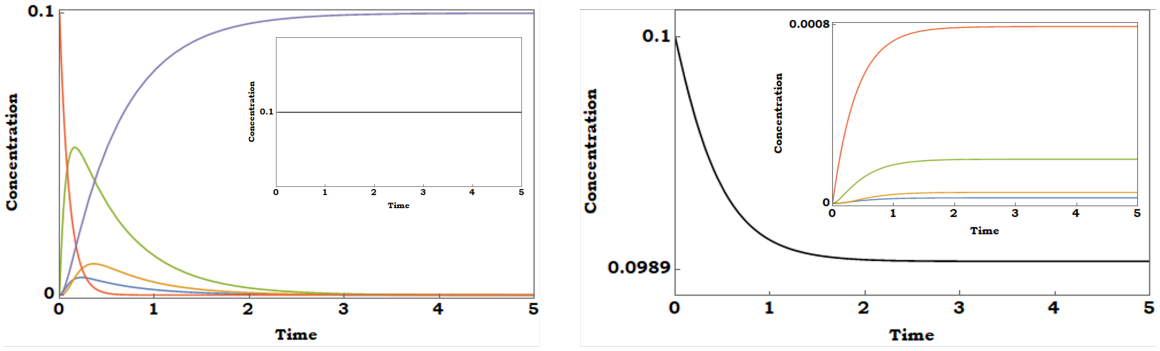
and the adjacency graph for the strongly connected graph is given by

$$\mathbf{A}^{\text{dir}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \\ 1/8 & 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 1/1000 \\ 0 & 0 & 1/8 & 0 & 0 \end{bmatrix}. \tag{6.5}$$

Figure 6.2 shows the propagation of toxic proteins on directed networks given in Figure 6.1. For the simulation, we choose a transport rate of $\rho = 20$ with an initial seed $c_4(0) = 0.1$ and $c_i(0) = 0$ for $i \neq 4$. Figure 6.2a shows that a seed initially starting at node 4 ends up at node 5. Dynamics at the other nodes increase and decrease as concentration moves from the source node to the sink node. The inset figure suggests that the process preserves the mass. Figure 6.2b shows that a seed distributes to the other nodes until they reach equilibrium state given by the Kernel of $\mathbf{L}^{\text{dir}}$. Topology of the graph in Figure 6.1b affects the distribution of protein concentration, which in this case a small amount of concentration distributes to nodes other than the seeding node shown in the inset figure. Although it is not shown, the process also preserves the mass. Overall, transport on directed graphs given by $\mathbf{L}^{\text{dir}}$ is strongly influenced by the topology of the network.

6.2.1 Transport on perturbed symmetric graph

Another way to describe the transport on directed networks is by perturbing a symmetric system with a small asymmetric factor. In this case, we perturb the undirected network with a small asymmetric network. We suppose $\mathcal{G}_0 = (\mathcal{V}, \mathcal{E}_0)$ be a symmetric



(a) (b)

Figure 6.2: Simulations on directed graphs in Figure 6.1 using $\rho = 20$: (a) Weakly connected graph and (b) Strongly connected graph.

graph with a collection of N nodes $\mathcal{V}$ representing regions of interest and edges $\mathcal{E}_0$ representing directed symmetric connections between these regions and $\mathcal{G}_1 = (\mathcal{V}, \mathcal{E}_1)$ be an asymmetric graph with the same collection of N nodes and edges $\mathcal{E}_1$ representing directed connections between the regions. If $\mathbf{A}_0$ and $\mathbf{A}_1$ respectively represent the adjacency matrices of graphs $\mathcal{G}_0$ and $\mathcal{G}_1$, then we define a directed adjacency matrix

$$\mathbf{A}^{\text{dir}} = \mathbf{A}_0 + \varepsilon \mathbf{A}_1, \tag{6.6}$$

where ε is an arbitrarily small value. We assume that $\mathbf{A}^{\text{dir}}$ to have positive elements to be biologically plausible since negative values mean instant loss of mass during the flow. In addition, we assume that $\mathbf{A}^{\text{dir}}$ is scaled so that $\max_{i,j}\{A_{0_{ij}}\} = 1$. The graph Laplacian of this topology also follows equation (6.2) and we have

$$\mathbf{L}^{\text{dir}} = \mathbf{L}_0 + \varepsilon \mathbf{L}_1, \tag{6.7}$$

where $\mathbf{L}_0$ and $\mathbf{L}_1$ denote the symmetric and asymmetric parts of the graph Laplacian, respectively. For a directed graph given as a perturbed symmetric system, we have the following proposition.

Proposition 2. *Consider the system* $(\mathbf{A} + \varepsilon \mathbf{B})\mathbf{x} = \lambda \mathbf{x}$, *with* $\boldsymbol{A}^T = \boldsymbol{A}$. *The complex eigenvalues* λ *can emerge when* $\boldsymbol{A}$ *has complete double eigenvalues, with eigenvectors* $\boldsymbol{u}_1$ *and* $\boldsymbol{u}_2$, *and satisfy* $(\boldsymbol{u}_2^T B \boldsymbol{u}_2 - \boldsymbol{u}_1^T B \boldsymbol{u}_1)^2 + 4(\boldsymbol{u}_2^T B \boldsymbol{u}_1)(\boldsymbol{u}_1^T B \boldsymbol{u}_2) < 0$.

Proof. Suppose λ_0 is a simple eigenvalue of $\mathbf{A}$ with eigenvector $\mathbf{x}_0$. Expanding $\mathbf{x} = \mathbf{x}_0 + \epsilon \mathbf{x}_1 + \ldots$ and $\lambda = \lambda_0 + \epsilon \lambda_1 + \ldots$ gives,

$$\mathbf{A}\mathbf{x}_0 - \lambda_0 \mathbf{x}_0 = 0, \tag{6.8}$$

$$\mathbf{A}\mathbf{x}_1 - \lambda_0 \mathbf{x}_1 = -\mathbf{B}\mathbf{x}_0 + \lambda_1 \mathbf{x}_0, \tag{6.9}$$

$$\mathbf{A}\mathbf{x}_2 - \lambda_0 \mathbf{x}_2 = -\mathbf{B}\mathbf{x}_1 + \lambda_1 \mathbf{x}_1 + \lambda_2 \mathbf{x}_0, \tag{6.10}$$

etc. Since $\mathbf{A}$ is symmetric, premultiplying (6.9) by $\mathbf{x}_0^T$ gives $\lambda_1 = \mathbf{x}_0^T\mathbf{B}\mathbf{x}_0/||\mathbf{x}_0||^2$. Similarly for (6.10), we have $\lambda_2 = (\mathbf{x}_0^T\mathbf{B}\mathbf{x}_1 - \lambda_1\mathbf{x}_0^T\mathbf{x}_1)/||\mathbf{x}_0||^2$. We can solve for λ_n at each order without any complex eigenvalues.

Now suppose that λ_0 is a double eigenvalue, with normalised eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$, chosen to be orthogonal. Then, $\mathbf{x}_0 = a\mathbf{u}_1 + b\mathbf{u}_2$. Then (6.9) gives $\mathbf{A}\mathbf{x}_1 - \lambda_0\mathbf{x}_1 = -\mathbf{B}(a\mathbf{u}_1 + b\mathbf{u}_2) + \lambda_1(a\mathbf{u}_1 + b\mathbf{u}_2)$ and the solvability conditions are $\lambda_1 = (a\mathbf{u}_1^T\mathbf{B}\mathbf{u}_1 + b\mathbf{u}_1^T\mathbf{B}\mathbf{u}_2)/a = (a\mathbf{u}_1^T\mathbf{B}\mathbf{u}_1 + b\mathbf{u}_2^T\mathbf{B}\mathbf{u}_2)/b$. Without loss of generality we can set $b = 1$. Then this is a quadratic equation for a, which may have complex roots, leading to complex λ_1. Specifically we have $a^2\mathbf{u}_2^T\mathbf{B}\mathbf{u}_1 + a(\mathbf{u}_2^T\mathbf{B}\mathbf{u}_2 - \mathbf{u}_1^T\mathbf{B}\mathbf{u}_1) - \mathbf{u}_1^T\mathbf{B}\mathbf{u}_2 = 0$ and there are complex roots if

$$(\mathbf{u}_2^T\mathbf{B}\mathbf{u}_2 - \mathbf{u}_1^T\mathbf{B}\mathbf{u}_1)^2 + 4(\mathbf{u}_2^T\mathbf{B}\mathbf{u}_1)(\mathbf{u}_1^T\mathbf{B}\mathbf{u}_2) < 0. \qquad \square \tag{6.11}$$

We have seen that the dynamics on directed networks are rich although their topology is simple. For instance, we may see oscillatory behaviour on a cycle graph associated with complex eigenvalue emergence from the system. Proposition 2 guarantees that a symmetric system that has double eigenvalues may have complex eigenvalues when it is perturbed. The consequence is that complex eigenvalues may be absent when the symmetric system does not have double eigenvalues. It is quite unlikely that we will observe oscillation in a weighted graph.

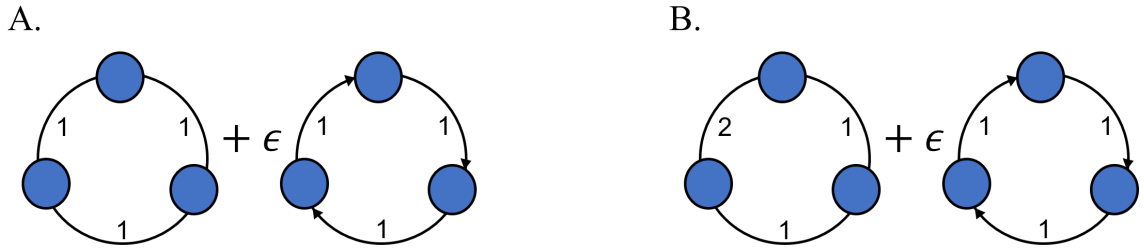


Figure 6.3: (a) Cycle-perturbed 3-node unweighted graph and (b) Cycle-perturbed 3-node weighted graph.

Figure 6.3 shows examples of a perturbed system that ε affects the system differently. The system in Figure 6.3A shows oscillation behaviour where the graph Laplacian nontrivial eigenvalues are $\lambda = 1/2\left(6 + 3\varepsilon \pm i\varepsilon\sqrt{3}\right)$ and the complex eigenvalues emerge for all ε. The perturbed weighted system in Figure 6.3B has its nontrivial eigenvalues $\lambda = 1/2\left(8 + 3\varepsilon \pm \sqrt{4 - 3\varepsilon^2}\right)$ and oscillation can emerge when $\varepsilon < -1.155$ or $\varepsilon > 1.155$. We notice that $\mathbf{A}^{\text{dir}}$ has negative elements in Figure 6.3B when $\varepsilon \leq -1.155$, which is not acceptable due to biological reasons.

6.2.2 Prion-like model on asymmetric network

We consider the invasion process on a network given by the Fisher-KPP equation (3.26) in Chapter 3 with the graph Laplacian given by (6.2) for directed graphs. The dynamics of the toxic protein propagation on a directed network is governed by

$$\begin{aligned} \frac{\mathrm{d}c_i}{\mathrm{d}t} &= -\rho \sum_{j=1}^{N} L_{ij}^{\mathrm{dir}} c_j + \alpha c_i(1-c_i), \qquad i = 1, ..., N, \\ c_i(0) &= q_i, \qquad\qquad\qquad\qquad\qquad\quad i = 1, ..., N. \end{aligned} \tag{6.12}$$

In the absence of transport ($\rho = 0$), the local dynamics give two steady states: the healthy state $\mathbf{c}^{(0)} = \mathbf{0}$ and the diseased state $\mathbf{c}^{(\infty)} = \mathbf{1}$. However, coupling the nonlinear kinetics with transport on a directed graph creates a new equilibrium state $\tilde{\mathbf{c}}^{(\infty)}$.

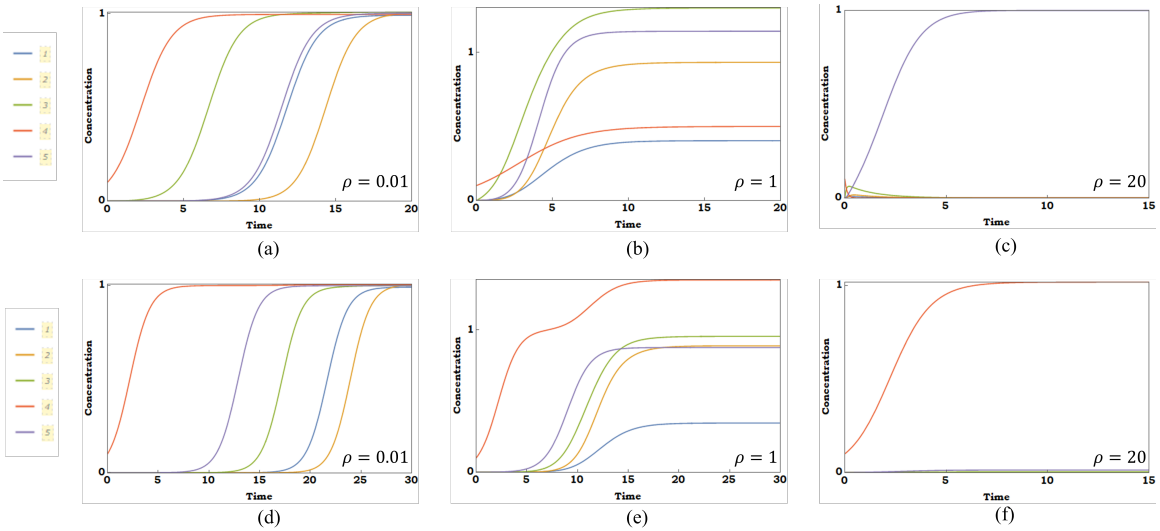


Figure 6.4: Fisher-KPP simulations on directed graphs in Figure 6.1 using $\alpha = 1$. Top: results on a weakly connected graph: (a) $\rho = 0.01$, (b) $\rho = 1$, and (c) $\rho = 20$. Bottom: results on a strongly connected graph: (d) $\rho = 0.01$, (e) $\rho = 1$, and (f) $\rho = 20$

Figure 6.4 shows toxic protein propagation using (6.12) on directed networks given in Figure 6.1. For the simulation, we choose a growth rate of $\alpha = 1$ and various mobility rates ρ. We refer $\rho/\alpha \ll 1$ as *the growth dominated*. While $\rho/\alpha \gg 1$, we refer to the condition as *the transport dominated*. Figures on the left panel (6.4a,d) depict the concentration evolution when parameters are in the growth-dominated regime. The behaviour strongly resembles the standard Fisher-KPP dynamics on a network, but each node grows towards a distinguished plateau. The behaviour in the

transport-dominated regime (Figure 6.4b,c,e,f) exhibits various dynamics depending on the transport rate ρ. The equilibrium of each node is clearly distinguished and the transport completely dominates the growth when ρ is large enough. The behaviour of (6.12) in the transport-dominated regime is different from the standard Fisher-KPP (3.26) in the same condition, where protein concentrations at all nodes evolve at the same rate showing one curve (see Chapter 3). This example suggests that the balancing process between growth and transport strongly affects the dynamics of (6.12) alongside the topology of networks.

6.3 Arrival times of asymmetric networks

In this section, we will explore the arrival times of toxic protein propagation on directed networks and derive arrival time estimates.

6.3.1 Linear arrival times

First, we linearise the system around the healthy state. We follow the same process to obtain linear arrival times as discussed in Chapter 4. The system in equation (6.12) reduces to the following linear form

$$\begin{aligned} \frac{\mathrm{d}c_i}{\mathrm{d}t} &= -\rho \sum_{j=1}^{N} L_{ij}^{\mathrm{dir}} c_j + \alpha c_i, \qquad i = 1, ..., N, \\ c_i(0) &= q_i, \qquad\qquad\qquad\qquad i = 1, ..., N. \end{aligned} \tag{6.13}$$

where ρ and α denote the transport and growth rate, respectively. Let $\mathbf{c}(t) = (c_1, ..., c_2)$, $\mathbf{q} = \mathbf{c}(0) = (q_1, ..., q_N)$, and $\mathbf{M} = -\rho \mathbf{L}^{\mathrm{dir}} + \alpha \mathbf{1}$, where $\mathbf{1}$ is the vector one. We can obtain a compact solution by integrating the system in terms of matrix exponentials. For a linearised system, the solution is given as follows

$$\mathbf{c}(t) = e^{t\mathbf{M}} \mathbf{q}. \tag{6.14}$$

Let us consider a complete set of eigenvectors $\mathbf{v}^{(i)} \in \mathbb{R}^N$ of $\mathbf{L}^{\mathrm{dir}}$ associated with eigenvalues $\lambda_i \in \mathbb{C}$, for $i = 1, ..., N$ so that

$$\mathbf{L}^{\mathrm{dir}} \mathbf{v}^{(i)} = \lambda_i \mathbf{v}^{(i)}. \tag{6.15}$$

Since the network is connected, we have zero eigenvalue $\Lambda_1 = 0$, which has Kernel of Span$\{\mathbf{v}^{(1)}\}$ but does not necessarily the span of vector one due to directed network

structures. Then, the linear solution can be written explicitly as

$$\mathbf{c} = \sum_{i=1}^{N} \left(\mathbf{q} \cdot \mathbf{v}^{(i)} e^{t\Lambda_i}\right) \mathbf{v}^{(i)}. \tag{6.16}$$

The general arrival time $\widehat{\tau}_j$ is given by the solution of the following equation

$$\mu = e^{\alpha \widehat{\tau}_j} \sum_{i=1}^{N} \left(\mathbf{q} \cdot \mathbf{v}^{(i)} e^{\widehat{\tau}_j \lambda_i}\right) \mathbf{v}_j^{(i)}. \tag{6.17}$$

and the canonical arrival time

$$\mu/\beta = e^{\alpha \widehat{\tau}_j} \sum_{i=1}^{N} \left(\mathbf{v}_j^{(i)} \mathbf{v}_s^{(i)} e^{\widehat{\tau}_j \lambda_i}\right). \tag{6.18}$$

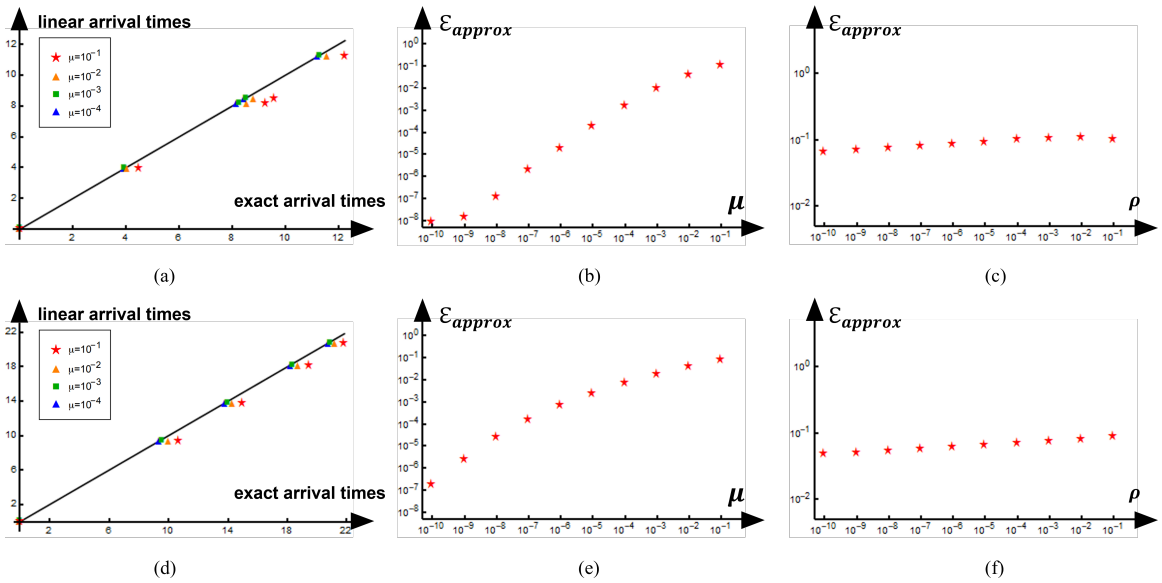


Figure 6.5: Simulations on directed graphs in Figure 6.1 using $\alpha = 1$. Top: results on a weakly connected graph. Bottom: results on a strongly connected graph. (a,d) linear time vs arrival time for $\rho = 0.01$, (b,e) error of linear times with varying $\mu = \beta$ for $\rho = 0.01$, and (c,f) error of linear times with varying ρ for $\mu = \beta = 0.1$.

We provide a comparison of the approximations with the numerical solutions for our 5-node networks to assess their performance. We define the error based on the exact and approximate arrival times as follows

$$\mathcal{E}_{\text{approx}} = \frac{\sum_{i=1}^{N} |\widehat{\tau}_i - \tau_i|}{\sum_{i=1}^{N} \tau_i}. \tag{6.19}$$

Figure 6.5 shows the performance of linear arrival times for the 5-node directed networks in Figure 6.1. We observe several general features. First, as we decrease $\beta = \mu$, the approximation given by the linear arrival times improves. This behavior is consistent with the dynamics for undirected graphs, where the solution is a credible approximation. The error analysis in Figure 6.5b,e shows that the approximation converges to the exact value. Notably, as shown in Figure 6.5c,f, the linear arrival times are independent of parameter changes. Second, the linear arrival times become a lower bound for the arrival times. This is due to the absence of nonlinear terms, which reduces the value of the concentration over time. As a result, the linear approach solutions are larger than the actual concentrations (excluding the zero arrival time at the seeding node).

We have a few remarks about the linear approach. First, it naturally combines the sum of all possible paths and walks in general, resembling a diffusive operator on undirected networks. Second, to solve N individual transcendental equations of the form (6.18), we need a good initial guess. Like the linear arrival times on undirected networks, we can use a Lambert method to provide this initial guess. Overall, the linear approach is useful for estimating arrival times on directed networks. It exhibits similar features to the undirected case and provides a robust and universal procedure for estimating arrival times.

6.3.2 Lambert times

We define a notion of time between nodes for this problem. It is the duration for an initial seed to propagate to a neighbouring node while ignoring other nodes in the network. Since we work with equal α, we expect to see a regular systematical front propagation from one node to another in the growth-dominated regime as depicted in Figure 6.4. We assume a similar process deriving the notion of times as discussed in Chapter 4 and evaluate the result accordingly.

Let us 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 (6.17) with surrogate graph Laplacian, corresponding to the two-node subnetwork of node i and j, given by

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

Using the eigenvectors of $\mathbf{L}_{ij}$, i.e. $\lambda_1 = 0$ and $\lambda_2 = A_{ij} + A_{ji}$, and (6.18) 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

$$e^{\alpha t}\left(1-e^{-\rho(A_{ij}+A_{ji})t}\right)\frac{A_{ij}}{A_{ij}+A_{ji}}=1. \tag{6.20}$$

Expanding for small ρ, we obtain

$$\rho A_{ij} t e^{\alpha t}=1. \tag{6.21}$$

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 $t=W_0(z)$):

$$t_{ij}=\frac{1}{\alpha}W_0\left(\frac{\alpha}{\rho A_{ij}}\right), \qquad i\neq j. \tag{6.22}$$

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

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

$$W_{ij}=\begin{cases}\displaystyle\sum_{k=0}^{n-1}t_{\gamma_k,\gamma_{k+1}}=\frac{1}{\alpha}\sum_{k=0}^{n-1}W_0\left(\frac{\alpha}{\rho A_{\gamma_k,\gamma_{k+1}}}\right), & i\neq j,\\ 0, & i=j,\end{cases} \tag{6.23}$$

which defines the *Lambert length matrix* $\mathbf{W}$. We find that this magnitude cannot be classified as a metric because the commutative property does not always hold due to asymmetric network structures.

The Lambert length 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

$$\frac{\mathrm{d}c_i}{\mathrm{d}t}=\alpha c_i\left(1-\frac{c_i}{c_{i,\infty}}\right), \quad c_i(0)=\beta,$$

satisfies $c(t_{ii})=\mu$ with equilibrium $c_{i,\infty}$. The solution to this problem is

$$t_{ii}=\frac{1}{\alpha}\log\left(\frac{(c_{i,\infty}-\beta)\mu}{(c_{i,\infty}-\mu)\beta}\right). \tag{6.24}$$

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

$$\tilde{\tau}_{ij} = t_{ii} + W_{ij}. \tag{6.25}$$

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, we assume that node i is chosen as the seeding location and we can compute the equilibrium $\tilde{c}_{j,\infty}$ numerically, with initial seed $c_i(0) = \mu$, then the dynamics at node j is given by

$$\frac{\mathrm{d}\tilde{c}_j}{\mathrm{d}t} = \alpha \tilde{c}_j \left(1 - \frac{\tilde{c}_j}{\tilde{c}_{j,\infty}}\right), \quad \tilde{c}_j(\tilde{\tau}_{ij}) = \mu, \tag{6.26}$$

where $\tilde{\tau}_{ij}$ is the Lambert arrival time (6.25). The *Lambert solution*

$$\tilde{c}_j = \frac{\tilde{c}_{j,\infty}\mu e^{\alpha t}}{\mu e^{\alpha t} - (\mu - \tilde{c}_{j,\infty})e^{\alpha \tilde{\tau}_{ij}}}, \quad j = 1, \ldots, N, \tag{6.27}$$

is therefore an approximation of (6.12).

We provide the Lambert edge length (6.22), Lambert length (6.23), and Lambert solution (6.27) for weakly and strongly connected graphs in Figure 6.1. The Lambert edge length and Lambert length matrices for the weakly connected graph are given in (6.28)

$$[t_{ij}] = \begin{bmatrix} \infty & 3.386 & \infty & \infty & \infty \\ \infty & \infty & 3.93 & \infty & \infty \\ 5.063 & \infty & \infty & \infty & 5.063 \\ \infty & \infty & 3.93 & \infty & 9.285 \\ \infty & \infty & \infty & \infty & \infty \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} 0. & 3.386 & 7.315 & \infty & 12.38 \\ 8.992 & 0. & 3.93 & \infty & 8.992 \\ 5.063 & 8.448 & 0. & \infty & 5.063 \\ 8.992 & 12.38 & 3.93 & 0. & 8.992 \\ \infty & \infty & \infty & \infty & 0. \end{bmatrix}, \tag{6.28}$$

and for the strongly connected graph in (6.29)

$$[t_{ij}] = \begin{bmatrix} \infty & 3.39 & \infty & \infty & \infty \\ \infty & \infty & 3.93 & \infty & \infty \\ 5.06 & \infty & \infty & 3.93 & \infty \\ \infty & \infty & \infty & \infty & 9.28 \\ \infty & \infty & 5.06 & \infty & \infty \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} 0. & 3.39 & 7.32 & 11.2 & 20.5 \\ 8.99 & 0. & 3.93 & 7.86 & 17.1 \\ 5.06 & 8.45 & 0. & 3.93 & 13.2 \\ 19.4 & 22.8 & 14.3 & 0. & 9.28 \\ 10.1 & 13.5 & 5.06 & 8.99 & 0. \end{bmatrix}, \tag{6.29}$$

where we use the symbol ∞ to characterise pairs of nodes that are not connected by an edge (implying infinite Lambert edge length). Figure 6.6 demonstrates the

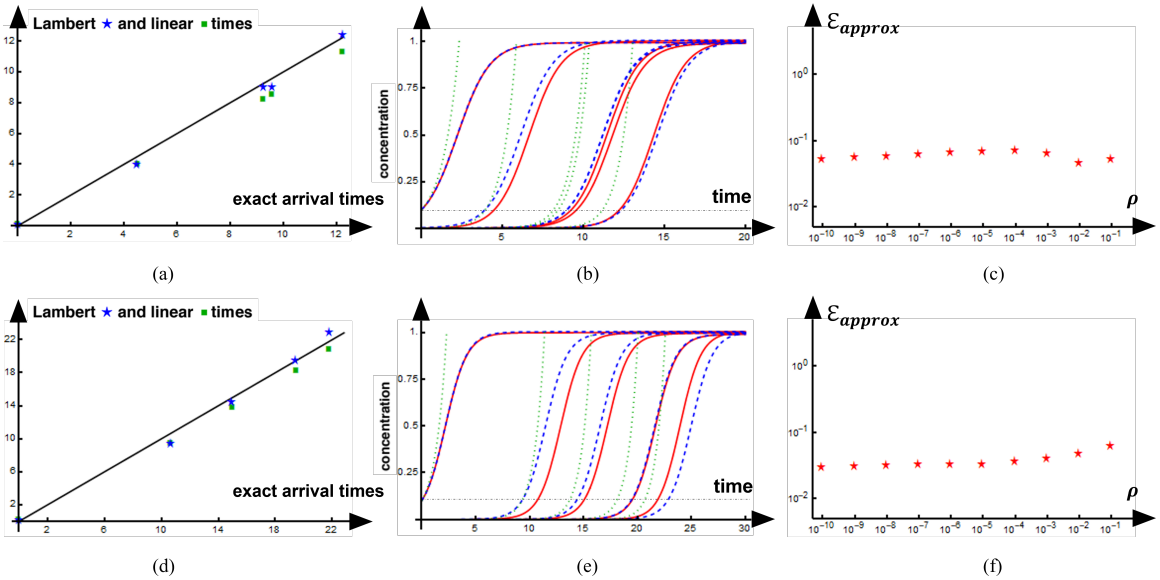


Figure 6.6: Simulations on directed graphs in Figure 6.1 using $\alpha = 1$ and $\rho = 0.01$. Top: results on a weakly connected graph. Bottom: results on a strongly connected graph. (a,d) Lambert time vs arrival time for $\rho = 0.01$, (b,e) exact (red), Lambert (dashed blue), and linear (dotted green) solutions, and (c,f) error of Lambert times with varying ρ for $\mu = \beta = 0.1$.

corresponding arrival times and Lambert solution. The Lambert method gives an improvement to the arrival time and solution. The approach is robust as the transport rate ρ varies as depicted in Figure 6.6c,f.

We note some remarks regarding the Lambert method on directed network case. First, we observe that this method provides a good estimate and is independent of initial concentrations. The method is simple since we can derive from two nodes and sum over the edge length over a path that connected a seed and a target node. Second, the Lambert length does include the diffusion role, but it cannot capture the advection effect. The derivation process of the length is the same as the Lambert distance for undirected network case despite they resemble different processes. Third, the Lambert length are neither an upper nor lower bound to the arrival times. It is because concentrations are diluted away from nodes before reaching the target node. Despite the path that strongly depends on the directions, the Lambert length can be underestimated or overestimated due to neglecting advection. Lastly, this method can be a first guess for computing the linear times despite its limitation. The features found in this method are similar to the Lambert distance for undirected network case in Chapter 4.

6.3.3 Nonlinear asymptotic times

The Lambert method suggests that propagation in the growth-dominated regime is similar to propagation on undirected networks, with the concentration spreading node by node to neighbouring nodes. However, in the case of directed networks, the path of propagation may differ from the underlying undirected topology, resulting in a different sequence of propagation. We present an alternative asymptotic method that captures important features of the solutions. For the canonical case, the *ansatz* is given as follows

$$\tilde{c}_i = \frac{\kappa_i K_i e^{\alpha t}}{\kappa_i + K_i \left(e^{\alpha t} - 1\right)}, \quad i = 1, \ldots, N. \tag{6.30}$$

where κ_i represents the new diseased state at node i due to asymmetric network configuration. We use a nonlinear version of the method of variation of constants, as introduced in [143],

$$K_i = A_i + \rho^{\gamma_i} B_i(t), \tag{6.31}$$

where γ_i represents the step distance between the seed node and node i. With this ansatz, we can iteratively solve 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 $\mathcal{O}(\rho^0)$, the solution at the seeding node with initial condition $c_s(0) = \beta$ is simply the original unperturbed solution

$$\tilde{c}_s = \frac{\kappa_s \beta e^{\alpha t}}{\kappa_s + \beta \left(e^{\alpha t} - 1\right)}, \tag{6.32}$$

and $A_i = 0$ for all $i \neq s$. The solutions for other nodes $\mathcal{O}(\rho^d)$ can then be obtained iteratively with nodes at distance d only depending on the solutions given by nodes at distance $d - 1$. More formally, if we define the matrix

$$M_{ij} = \begin{cases} \dfrac{L_{ij}}{\kappa_i} & \text{if } \gamma_i - \gamma_j = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{6.33}$$

and the variables

$$Z_i = \begin{cases} B_i, & \text{if } i \neq s \\ e^{-\alpha t} \tilde{c}_s & \text{if } i = s, \end{cases} \tag{6.34}$$

then, these new variables are solutions to the system

$$\frac{\mathrm{d}Z_i}{\mathrm{d}t} = \sum_{j=1}^{N} M_{ij} Z_j, \qquad Z_i(0) = 0, \qquad i \neq s. \tag{6.35}$$

To order $\mathcal{O}(\rho^1)$, we obtain

$$B_i(t) = \frac{\beta\kappa_s L_{is}}{\alpha(\beta-\kappa_s)\kappa_i}\left(\alpha t - \log\left(\beta\left(e^{\alpha t}-1\right)+\kappa_s\right)+\log\left(\kappa_s\right)\right) \tag{6.36}$$

and the solution of (6.35) is given in the polylogarithm function Li_d where d is the distance to the seeding node.

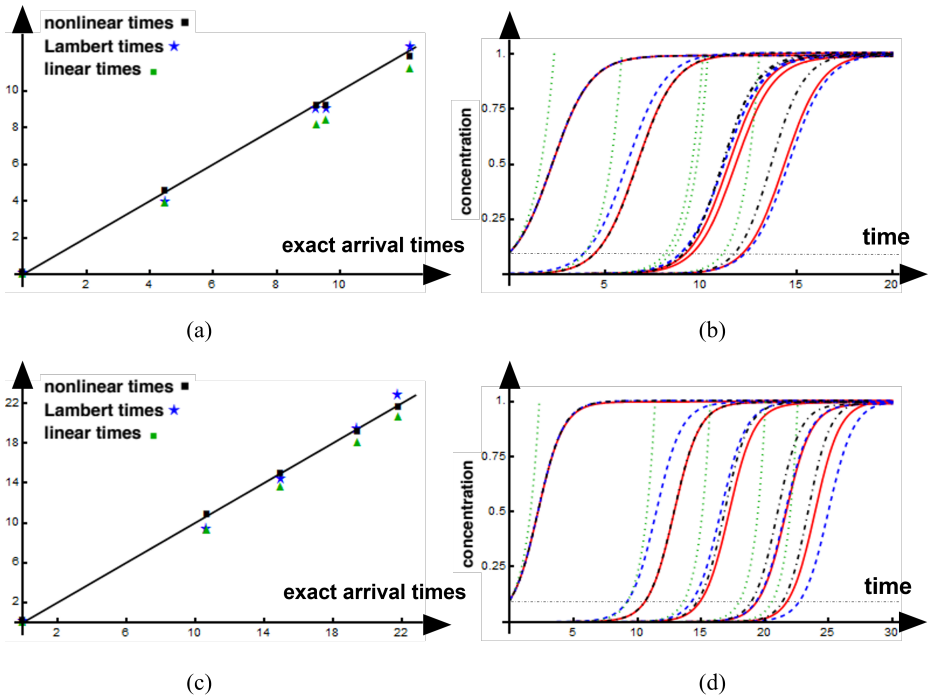


Figure 6.7: Simulations on directed graphs in Figure 6.1 using $\alpha = 1$ and $\rho = 0.01$. Top: results on a weakly connected graph. Bottom: results on a strongly connected graph. (a,c) Nonlinear time vs arrival time for $\rho = 0.01$, (b,d) exact (red), nonlinear (dash-dotted black), Lambert (dashed blue), and linear (dotted green) solutions.

Figure 6.7 demonstrates the arrival times and solution of the Fisher-KPP (6.12) on the directed networks in Figure 6.1. We assume that we can compute the new equilibrium and substitute the solution of (6.35) to (6.30), the *nonlinear asymptotic solutions* satisfy the initial values and the asymptotic behaviour for a long time. This method works for both weakly and strongly connected graphs. We compute the errors of arrival times in Figure 6.7a,c and find that the errors decrease as we move from linear times to Lambert times to nonlinear times: 1.08×10^{-1}, 4.44×10^{-2}, and 2.84×10^{-2} for the weakly connected graph and 7.91×10^{-2}, 4.61×10^{-2}, and 1.54×10^{-2} for the strongly connected graph. Notably, the nonlinear method significantly improves upon both the Lambert times and Lambert solutions.

We note several features of the nonlinear method. In general, this method is similar to the undirected graph, but different in formulation. First, the solution is given by a polylogarithmic function Li_d where d is the distance to the seeding node. Given κ_i, the solution satisfies the initial condition and asymptotic behaviour for long time. Second, the method includes the saturating effect and, implicitly, the advection effect. In the ansatz, we need to input κ_i which is the asymptotic behaviour of the solution due to balancing between the transport and kinetics. In addition, κ_i is included in the step-by-step connectivity. Third, this approach naturally combines all contributions of paths of length d. However, they do not include paths of different lengths and, thus, do not address the problem of dilution away from a node. Despite its limitations, the method improves our estimates and captures the spatiotemporal evolution on directed graphs.

6.4 Application to neurodegenerative disease

In this section, we apply methods to estimate the arrival time to the study of propagation of toxic protein transported on a structural connectome and explore our models with respect to staging. We assume that the directional bias comes from the degree of nodes and local capacity data. We use the scale-33 structural connectome that has 83 nodes. Each node represents a brain region. To produce staging, we use the six Braak stages where the corresponding regions of each stage are based on the Braak staging given by ADNI.

Brain regions corresponding to Braak stages are depicted in Figure 6.8. The six stages are a sequence of τP hierarchical propagation in the brain [18, 17]. The starting location is the entorhinal cortex associated with Braak stage I. The expected progressive pattern of dynamics follows the Braak staging. For the scale-33 brain connectome, the entorhinal cortex corresponds to nodes 27 and 68.

The local capacity data are obtained with the same process in Chapter 5. From the data, we infer the local parameters for each node representing regional baseline values for a healthy state and carrying capacities for a late-stage AD state. The parameters used in this study include information about the PET measurement process and AD, such as off-target binding and regional variations in production dynamics. The Gaussian mixture modelling approach developed in [180] is used to obtain these parameters. To estimate the fixed parameters for $c_{i,0}$ and $c_{i,1}$ representing the minimum and maximum τP SUVR scores, respectively, we fit a two-component Gaussian mixture model to population-level data of regional SUVR. For regions where a reliable

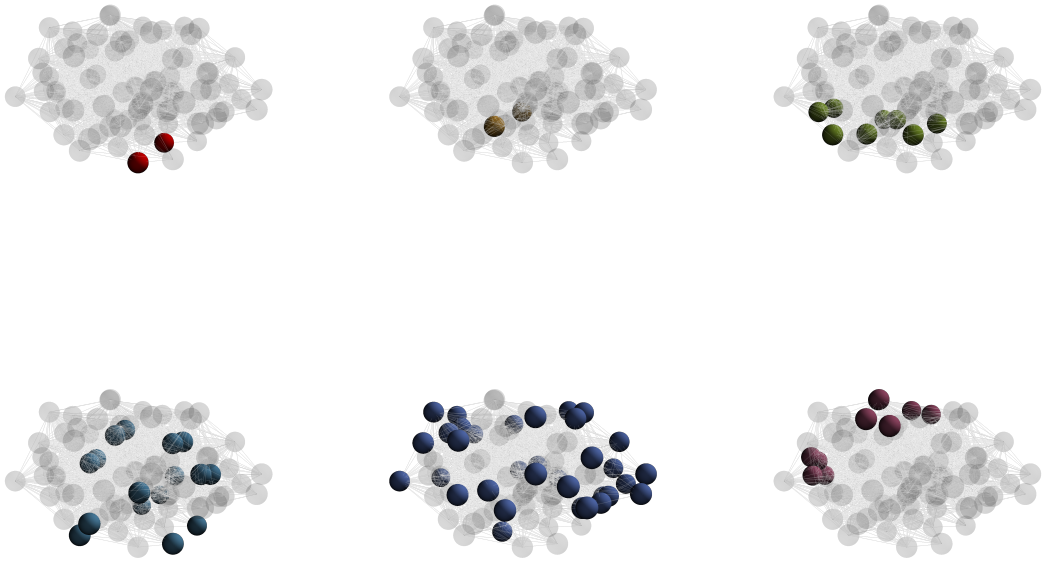


Figure 6.8: Braak regions of interest according to ADNI. Top: Braak I (left), Braak II (middle), Braak III (right). Bottom: Braak IV (left), Braak V (middle), Braak VI (right).

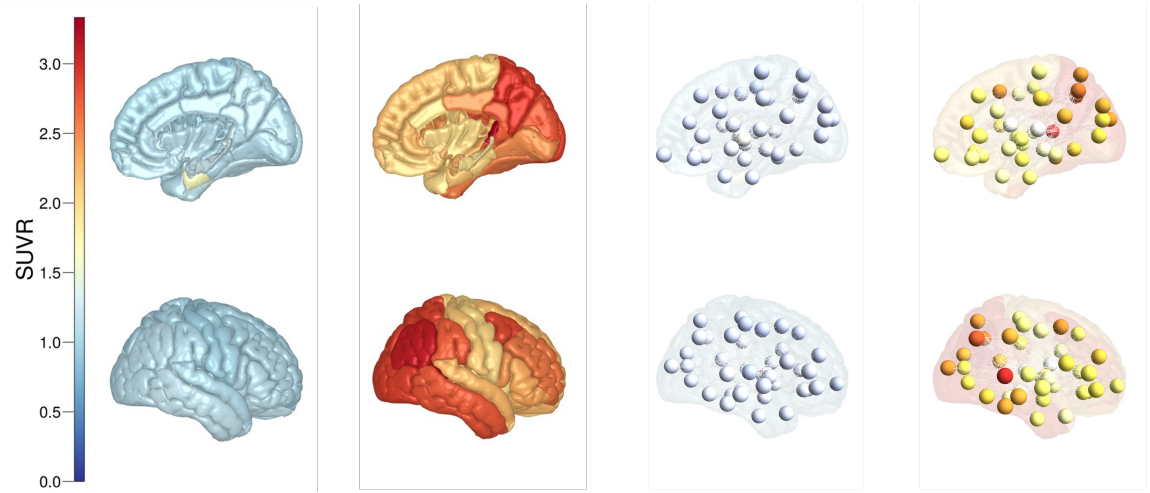


Figure 6.9: Local τP SUVR scores across the brain. Brains on the first two columns (left) respectively show the minimum and maximum level of τP SUVR scores. The last two columns (right) show their associated networks.

measure of τP SUVR can be obtained, we expect to see two separate distributions, a τP$-$ distribution capturing the expected τP load in a given region, and a τP+ distri-

bution describing the pathological τP load [180]. Using the fitted Gaussian mixture models, we approximate $c_{i,0}$ as the mean of the τP$-$ distributions and $c_{i,1}$ as the 99th percentile of the τP+ distributions. Due to off-target binding, it is not possible to obtain τP PET signal for subcortical regions [32, 158]. Therefore, these regions are excluded from our model, leaving a total of 72 regions. The multi-tau cohort of AV1451 PET data detailed in [180] is used for ADNI. Figure 5.7 shows the regional minimum and maximum capacities in the brain. If we define the regional capacity difference, the caudate has the smallest values while the middle temporal has the largest values. We define regional carrying capacity $\mathcal{K}_i = c_{i,1} - c_{i,0}$.

We derive the directed structural connectomes either by multiplying the inverse of the node degree matrix with the adjacency matrix or by multiplying the inverse of the regional carrying capacity matrix with the adjacency matrix. If $\mathbf{A}, \mathbf{D}$, and $\mathbf{K}$ denote the adjacency matrix of the structural connectome, the node degree matrix of $\mathbf{A}$, and the diagonal matrix which its elements $K_{ii} = \mathcal{K}_i$, respectively, then we define the directed network $\mathbf{A}^{\text{dir}}$ to be

$$\mathbf{A}^{\text{dir}} = \mathbf{D}^{-1}\mathbf{A}, \tag{6.37}$$

for the asymmetric network from the degree of node and

$$\mathbf{A}^{\text{dir}} = \mathbf{K}^{-1}\mathbf{A}, \tag{6.38}$$

for the asymmetric network from the local capacity difference.

6.4.1 Tau protein propagation

We apply the methods that we derive for arrival time estimation on the directed network problem. In this subsection, we demonstrate the comparison between methods and the solution of the nonlinear asymptotic approach to the τP propagation. Instead of depicting the evolution of each node, we compute the average dynamics of Braak regions. The seeding location is the entorhinal cortex represented by nodes 27 and 68. The initial values are $c_{27}(0) = c_{68}(0) = 0.1$ and zero otherwise.

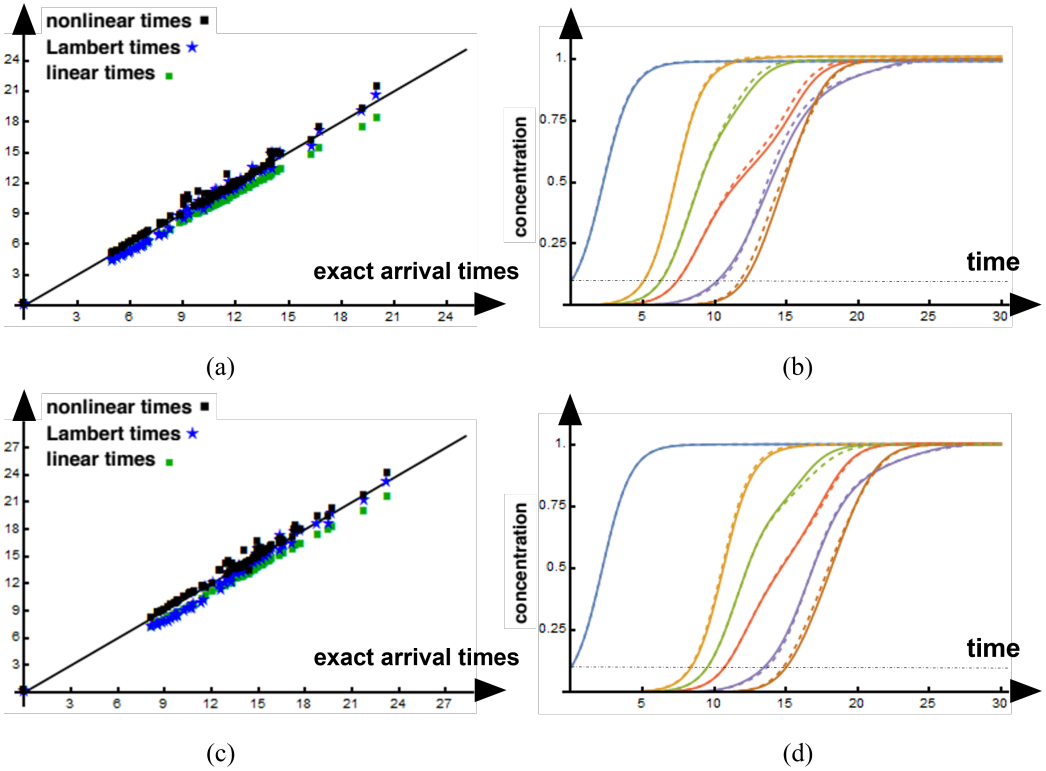


Figure 6.10: Simulations on asymmetric networks using nonlinear method. Top: Simulation on the asymmetric network from the degree of node 6.37. Bottom: Simulation on the asymmetric network from the local capacity difference 6.38. (a,c) Nonlinear time vs arrival time for $\rho = 0.01$, (b,d) exact (bold), nonlinear (dashed).

Figure 6.10 shows the application of methods on (6.12). The directed connectome is obtained by left-multiplying the weighted adjacency matrix with a diagonal matrix. This procedure follows the way to define a retrograde transport in the connectome [131]. The first directed adjacency matrix given by (6.37) allows the concentration to flow from higher-degree nodes to lower-degree nodes. The first directed adjacency matrix given by (6.38) allows the concentration to flow from a region with higher capacity to regions with lower capacity. We use $\alpha = 1$ and $\varepsilon = 0.01$. The errors decrease as we move from the linear times, Lambert times, to nonlinear times: 1.09×10^{-1}, 4.74×10^{-2}, and 2.84×10^{-2} for the degree of node and 9.82×10^{-2}, 5.1×10^{-2}, and 1.93×10^{-2} for the local capacity difference. The nonlinear solution provides a decent approximation.

Our results show that the nonlinear asymptotic method provides a good, and best overall, estimate for both arrival times and dynamics. In particular, the approach perfectly captures the average dynamics of asymmetric networks. The method, however, needs a crucial input that depends on the numerical simulation due to the emergence

of new asymptotic values.

6.4.2 Staging of tau protein

We have applied each approximation method to directed network systems. We are interested in exploring the influence of directional bias in the system towards the resulting patterns. We assess each model systematically through staging analysis in the growth-dominated regime ($\rho/\alpha \ll 1$) as introduced in Chapter 3.

To understand the influence of parameter changes on staging, we perturb the symmetric adjacency matrix according to (6.6). Therefore, the symmetric and asymmetric parts of the adjacency matrix are given by

$$\begin{aligned} \mathbf{A}_0 &= \frac{1}{2}\left(\mathbf{A}^{\text{dir}} + \left(\mathbf{A}^{\text{dir}}\right)^T\right), \\ \mathbf{A}_1 &= \frac{1}{2}\left(\mathbf{A}^{\text{dir}} - \left(\mathbf{A}^{\text{dir}}\right)^T)\right), \end{aligned} \tag{6.39}$$

where ε is a small tuning parameter in the interval of $[-1, 1]$. The value ε is to ensure positive weights. The value of $\varepsilon = 1$ defines the retrograde flow while $\varepsilon = -1$ defines the anterograde flow. The graph Laplacian is then given by (6.2).

We note that the asymptotic behaviour on the long time of directed network systems can be different from the dynamics on undirected networks which the asymptotic behaviour is the Kernel of the graph Laplacian. We introduce a new magnitude in the range of 0 and 1 out of J regions $\Omega_1, ..., \Omega_J$. In each of these regions, we define

$$C_j(t) = \frac{1}{\sigma_{\Omega_j}} \sum_{i \in \mathcal{V}_j} c_i(t), \quad j = 1, ..., J, \tag{6.40}$$

where

$$\sigma_{\Omega_j} = \sum_{i \in \mathcal{V}_j} c_{i,\infty},$$

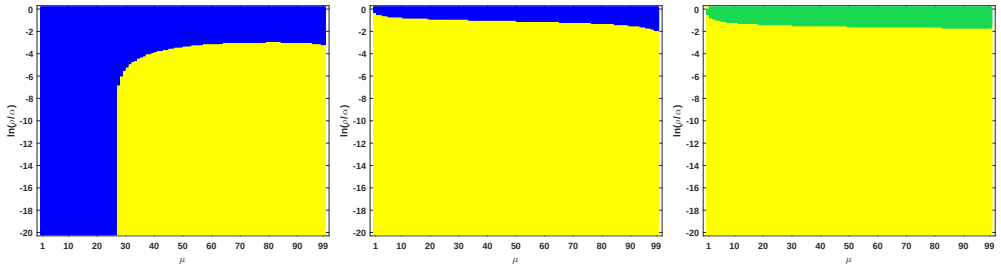
where $\mathcal{V}_j$ is the collection of nodes defining region Ω_j with N_j nodes and $c_{i,\infty}$ represents the nontrivial asymptotic concentration at node i. The average value (6.40) accommodates the new asymptotic without forcing each node to evolve between 0 and 1. In our simulations, the first region Ω_1 is the bilateral entorhinal cortex with initial conditions $c_i(0) = 0.005/N_1$ for $i \in \mathcal{V}_1$ and zero otherwise. We determine two staging sequences of the first three stages and the last three stages. Using this staging, we can see clearly how directional bias alters the patterns resulting from a directed network system. For a pair of parameter ρ/α and threshold value, we assign a colour label.

The assigned colours are shown in Table 6.1. The expected canonical Braak staging of τP is given by progressive sequences I $\rightarrow$ II $\rightarrow$ III for the first three Braak stages and IV $\rightarrow$ V $\rightarrow$ VI for the last three stages. The yellow colour indicates the canonical staging sequence. We summarise all computed colours in our simulations into braid surfaces.

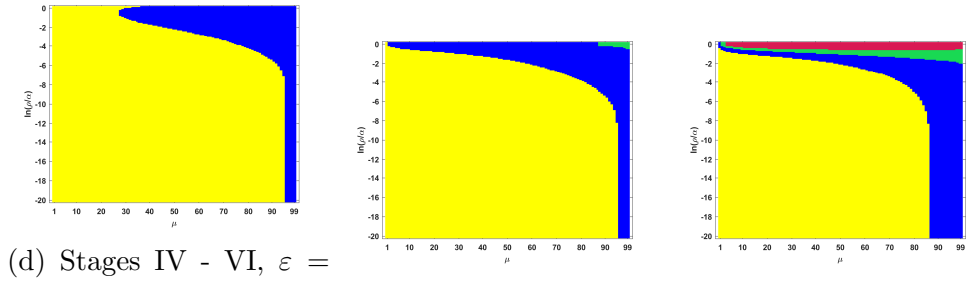
All observed computational staging patterns				
Braak regions I, II, and III		I $\rightarrow$ II $\rightarrow$ III *		I $\rightarrow$ III $\rightarrow$ II
		Beginning with stage II		Beginning with stage III
Braak regions IV, V, and VI		IV $\rightarrow$ V $\rightarrow$ VI*		IV $\rightarrow$ VI $\rightarrow$ V
		V $\rightarrow$ IV $\rightarrow$ VI		V $\rightarrow$ VI $\rightarrow$ IV
		VI $\rightarrow$ IV $\rightarrow$ V		VI $\rightarrow$ V $\rightarrow$ IV

Table 6.1: Color coding for braid surfaces of τP propagation. * Progressive computational Braak staging.

Figure 6.11 shows staging patterns when using (6.37) to create a directed weighted adjacency matrix while Figure 6.12 shows staging patterns when using (6.38). In general, the staging patterns shown by the two procedures demonstrate a similar behaviour. The resulting patterns when we change ε show small alterations and are rather consistent compared to the patterns of undirected connections where $\varepsilon = 0$. Notable differences between patterns resulting from two procedures are shown for staging at higher ρ/α values near the diffusion-dominated regime and the last three-stage staging patterns. We also note that higher anterograde transport on the connectome influences the concentration to accumulate in region III prior to region II in the growth-dominated regime before the staging is stable at the canonical sequence when the degree of node is applied. Further results with various ε values are shown in the Appendix 7.2.2. Our results suggest that, in the existence of directions, the overall dynamics of Braak staging show little change in the growth-dominated regime where we believe τP propagates under this condition. The symmetric connectivities have become a strong foundation and regulate the flow of τP across the brain.

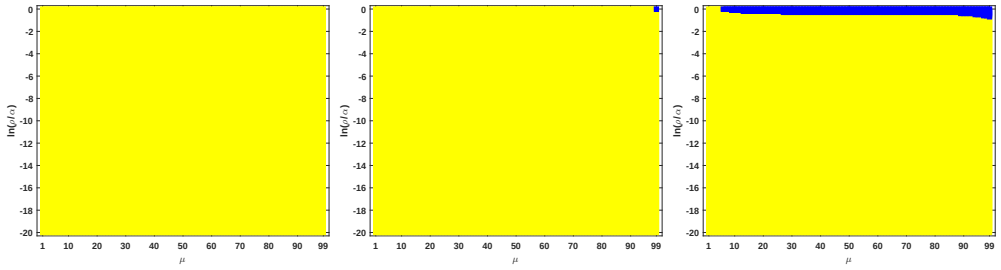


(a) Stages I - III, $\varepsilon = -1$ (b) Stages I - III, $\varepsilon = 0$ (c) Stages I - III, $\varepsilon = 1$

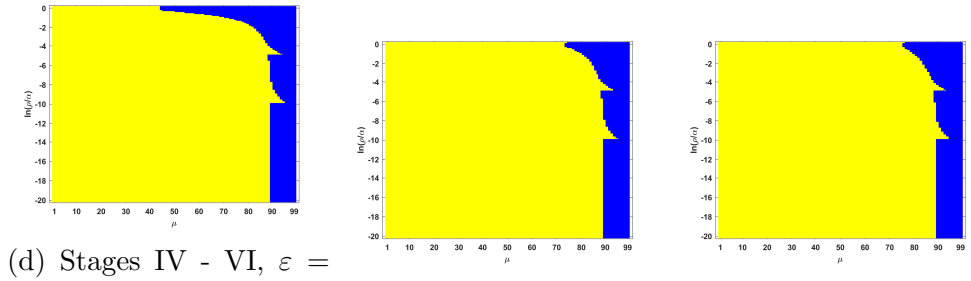


(d) Stages IV - VI, $\varepsilon = -1$ (e) Stages IV - VI, $\varepsilon = 0$ (f) Stages IV - VI, $\varepsilon = 1$

Figure 6.11: Observed computational (deterministic) connectome τP seed staging; Stages I - III (top) and stages IV - VI (bottom). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (6.12) with $\mathbf{L}^{\text{dir}}(6.7)$ from $\mathbf{A}^{\text{dir}}$ (6.37).



(a) Stages I - III, $\varepsilon = -1$ (b) Stages I - III, $\varepsilon = 0$ (c) Stages I - III, $\varepsilon = 1$



(d) Stages IV - VI, $\varepsilon = -1$ (e) Stages IV - VI, $\varepsilon = 0$ (f) Stages IV - VI, $\varepsilon = 1$

Figure 6.12: Observed computational (deterministic) connectome τP seed staging; Stages I - III (top) and stages IV - VI (bottom). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (6.12) with $\mathbf{L}^{\text{dir}}(6.7)$ from $\mathbf{A}^{\text{dir}}$ (6.38).

Our results show that directional bias in the system influences the landscape of observable staging patterns. In particular, pattern alterations occur when parameter values change considerably. The most apparent is the influence of strong directional bias when $\varepsilon = -1$ or $\varepsilon = 1$.

6.5 Concluding remark

In this chapter, we explore the problem of directed network systems that arise from studying the propagation of τP on the brain connectome. We aim to find any possible models that account for the directional links in the brain network. Building on Chapters 3 and 4, we examine how the directional bias affects the overall dynamics and estimates of the system.

We consider the role of growth and transport in our toxic τP propagation mathematical model. In this chapter, we assume a directional bias transport on the brain network. The transport is given by the graph Laplacian for directed networks, which still preserves total mass but does not satisfy Fick's condition. The directionality inclusion in the system alters the asymptotic behaviour for graph Laplacian and

the full system of the toxic τP propagation. The diseased state is a nontrivial heterogeneous equilibrium. Staging analysis suggests that the directionality affects the resulting staging patterns, particularly when the directional bias is strong. The study of arrival times suggests that the propagation of toxic τPs on directed networks is similar to the transport on undirected networks. This feature is observed through the three estimates for the arrival times to understand the underlying mechanism of propagation on directed networks.

The first approach is to linearise the full system around the healthy state. We then obtain the solution of protein concentration and solve the transcendental equation to obtain linear arrival times. The method is universal for the transport on directed networks.

The second approach is to derive an estimate between two adjacent nodes, called the *Lambert edge length* or *Lambert edge distance*, and take the sum over a path. The path is likely the shortest path since the sum of estimates gives the smallest value and dominates the propagation. The derived estimate is the same for undirected networks. However, the magnitude is not a distance metric. Despite, its limitation, the Lambert method is computationally cheap and independent of initial values.

The last method is an improvement on the Lambert solution by including nonlinearity in the solution. The approach takes into account a natural small parameter which is similar to the approach for undirected networks. The emergence of new asymptotic values is also considered in the solution. The solution is in the polylogarithm function that considers multiple paths with the same length. The process to compute the nonlinear method is similar to the method for undirected networks in Chapter 4. However, we still need to compute numerically the equilibrium for the full system. This method shows a promising estimate despite its complexity and long expression.

Our study on staging and arrival time problems allows us to explore the influence of directional bias on the propagation process from a seed to its neighbouring nodes. To some extent, the directionality of networks influences the resulting staging sequences. The dominating factor of invasion depends on the shortest path as shown by the Lambert method and nonlinear asymptotic approach. Our exploration of methods for estimating arrival times shows that these methods can accommodate the Fisher-Kolmogorov-Petrovsky-Piskunov dynamical process on directed networks. Overall, staging and arrival time problems are essential in understanding the complex coupled process from both computational and analytical points of view.

Chapter 7

Epilogue

7.1 Conclusion

This thesis explored how τP propagates on the brain connectome and how it relates to its hierarchical staging pattern and the time of arrival of the τP seed in different brain regions.

In Chapter 2, we reviewed the evidence and mechanisms of prion-like neurodegenerative diseases and the mathematical models that have been used to study them. We also formulated three key problems that we addressed in the subsequent chapters.

Chapter 3 explored the model selection problem for τP hierarchical staging in Alzheimer's disease. We provided a general formulation for τP propagation in the brain subjected to physical diffusion on networks. A simple one species τP reaction-diffusion on a network coupled with a damage measure was derived and plots for τP and damage staging assessment, called braid diagram and braid surface, were introduced. We provided statistical analysis using τP PET SUVR scores to validate our finding of possible staging patterns.

Chapter 4 analysed the invasion process of τP on networks based on the two propagation regimes observed in Chapter 3. This study emphasised the time of τP seed arrival to understand the transport process on the network. We derived three estimates of the time of arrival where the linearised system becomes a foundation: the linear arrival time, the Lambert arrival time, and the nonlinear arrival time. The best estimate was based on a new asymptotic method and provided an approximation of the solutions for all time.

Chapter 5 presented a follow-up study of τP propagation on networks by considering heterogeneity in the kinetic term. We followed a similar procedure as Chapter 4 to derive the estimates of the arrival time. The linearised system is a critical foundation to derive the approximation and taking into account the nonlinear behaviour provides

a better estimate. In this chapter, we also considered the staging for heterogeneous systems with small parameter changes.

Chapter 6 presented another follow-up study of τP propagation on a network considering directed networks. The procedure to derive the estimates of the time of arrival is similar to Chapter 4. We introduced the same notion of linear and Lambert estimate for invasion dynamics on directed networks. The best estimate includes nonlinear behaviour and additional asymptotic behaviour of the system. In this chapter, we also considered the staging for small perturbations to symmetric networks.

To conclude, the study of τP propagation on networks with respect to staging and time of arrival enables one to investigate spatiotemporal evolution in the brain. Braid diagram and braid surface are promising tools for model selection and qualitative assessment of regional sequence permutation while arrival time estimates provide a meaningful approximation in order to understand invasion processes.

7.2 Future Work

In this thesis, we have studied the progression of biomarkers in the human brain, specifically the in vivo dynamic of toxic tau protein in Alzheimer's disease. Though our results introduce some important concepts such as staging and arrival time, there are some open questions that need immediate follow-up. Some possible directions are:

- Studying the propagation of tau on directed networks in the diffusion-dominated regime ($\rho/\alpha \gg 1$), which may show different dynamics and equilibrium points than the growth-dominated regime ($\rho/\alpha \ll 1$).
- Applying the arrival time estimates to computational problems such as parameter estimation, wave speed computation, and network coarse-graining.
- Studying the relation between arrival time and neuron damage.
- Arrival time comparison between Fisher-KPP and Heterodimer model.
- Understanding staging information loss between different connectome resolutions.

Finally, we list two other possible research extensions that we find interesting.

7.2.1 Amyloid-tau protein propagation to study regional vulnerability

Neurodegenerative diseases affect different types of neurons in different brain regions, resulting in distinct trajectories of the disease. For example, Alzheimer's disease is associated with tau pathology, which preferentially targets excitatory neurons at the early stage and leads to significant loss of excitatory neurons at the later stage [18, 56, 146]. Tau pathology also shows different patterns in different subtypes of Alzheimer's disease, which have different cognitive profiles and outcomes [181]. The aim is to understand how regional vulnerability is reflected in the evolution of τP biomarkers in the brain.

One way to study regional vulnerability is to use mathematical models for the propagation of Aβ and τP in the brain. A simple and comprehensive model that incorporates both proteins is the conspiracy model by Thompson et al [172]. This model considers the interaction of both proteins and how they cause primary or secondary tauopathy. It shows that toxic Aβ can enhance the effects of toxic τP in both cases. It also shows that the distribution of damage depends on the seeding location of each protein. This suggests different types of neurodegeneration in AD. A simplified version of the model based on Fisher-KPP equations can capture the qualitative features of primary and secondary tauopathies [104]. A model that includes both Aβ and τP can provide insights into disease localisation and regional vulnerability in neurodegenerative diseases.

A possible extension of our study is to investigate regional vulnerability in neurodegenerative disease using staging and arrival time methods. We modelled τP propagation in the brain with Fisher-KPP and suggested braid diagram and braid surface to replicate and evaluate staging. We aim to identify a spatiotemporal pattern and visualise it as a braid diagram or surface. We also want to map regional vulnerability based on toxic protein propagation and interaction, such as Aβ and τP. We hope that this extension can reveal new insights into brain damage caused by protein-protein interaction.

7.2.2 Invasion-dynamic-based community detection

We can also explore community detection based on the invasion process. Let us define t_i be the arrival time at node i, when a quantity of interest $c_i(t_i) = T$ reaches a threshold T. Let us define $\mathbf{\Delta T}$ to be a time difference matrix with elements $\Delta t_{ij} = |t_i - t_j|$. Figure 7.1 demonstrates Fisher-KPP simulations using $\alpha = 1$ with $\log(\rho) =$

-30 or $\log(\rho) = -3$ on the Zachary's Karate Club network. For each simulation, we record the arrival times of nodes at a threshold value and compute $\mathbf{\Delta T}$. We aim to identify and visualise a spatiotemporal pattern as a braid diagram or surface. We also want to detect communities based on toxic protein propagation and interaction.

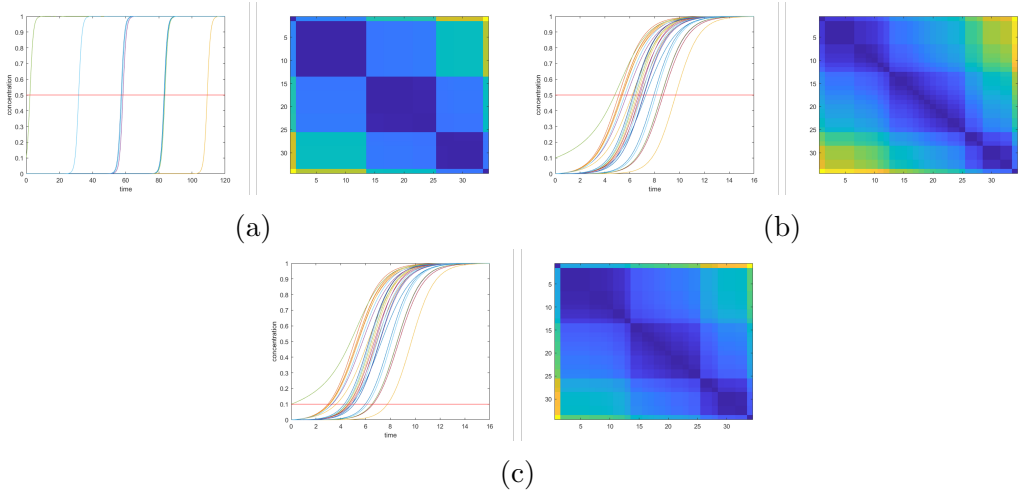
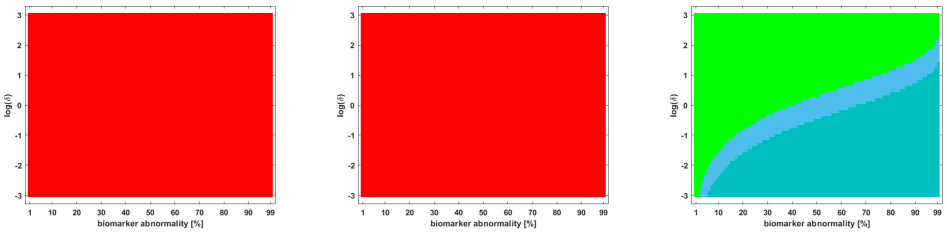


Figure 7.1: Arising communities under invasion process. (a) Fisher-KPP simulation using $\alpha = 1$ and $\log(\rho) = -30$ and time difference between nodes at 0.5 threshold, (b) Fisher-KPP simulation using $\alpha = 1$ and $\log(\rho) = -3$ and time difference between nodes at 0.5 threshold, and (c) Fisher-KPP simulation using $\alpha = 1$ and $\log(\rho) = -3$ and time difference between nodes at 0.1 threshold.

We propose to use invasion dynamics to detect communities in networks, as shown in Figure 7.1. We compute the time difference matrix $\mathbf{\Delta T}$ from the arrival times of nodes at a threshold value. The matrix suggests different numbers of communities for different diffusion rates. This resembles the multiscale community detection method by Delvenne et al [42], which uses Markovian diffusion flows to identify relevant communities at different time scales. We face some challenges in defining graph partitions based on $\mathbf{\Delta T}$, such as choosing a suitable threshold and dealing with the influence of seeding locations and numbers. We also want to compare our results with other well-known methods and explore the possibility of retrieving them from our approach. This may lead to new insights on community detection.

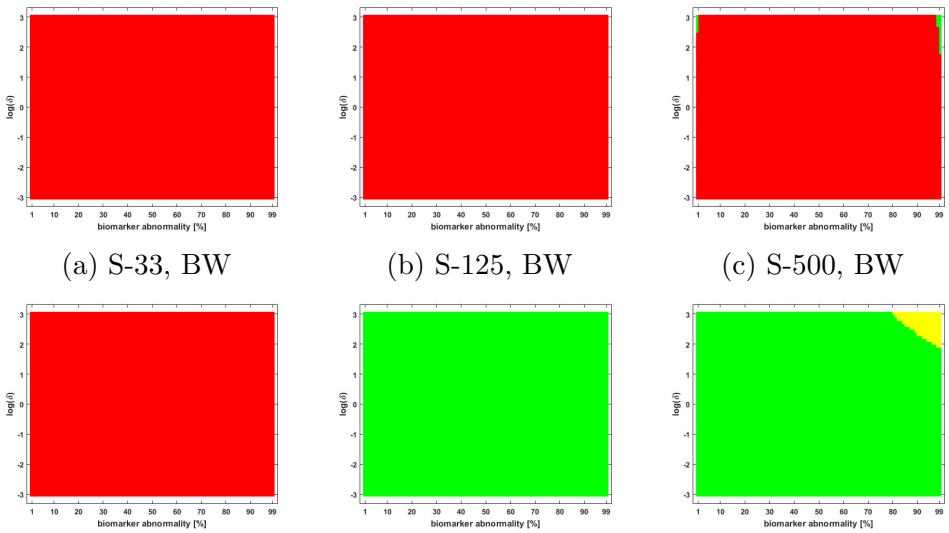
Appendix A
Additional Staging Results of Chapter 3

Additional results, deterministic streamlined connectome staging



(a) S-33, LW (b) S-125, LW (c) S-500, LW

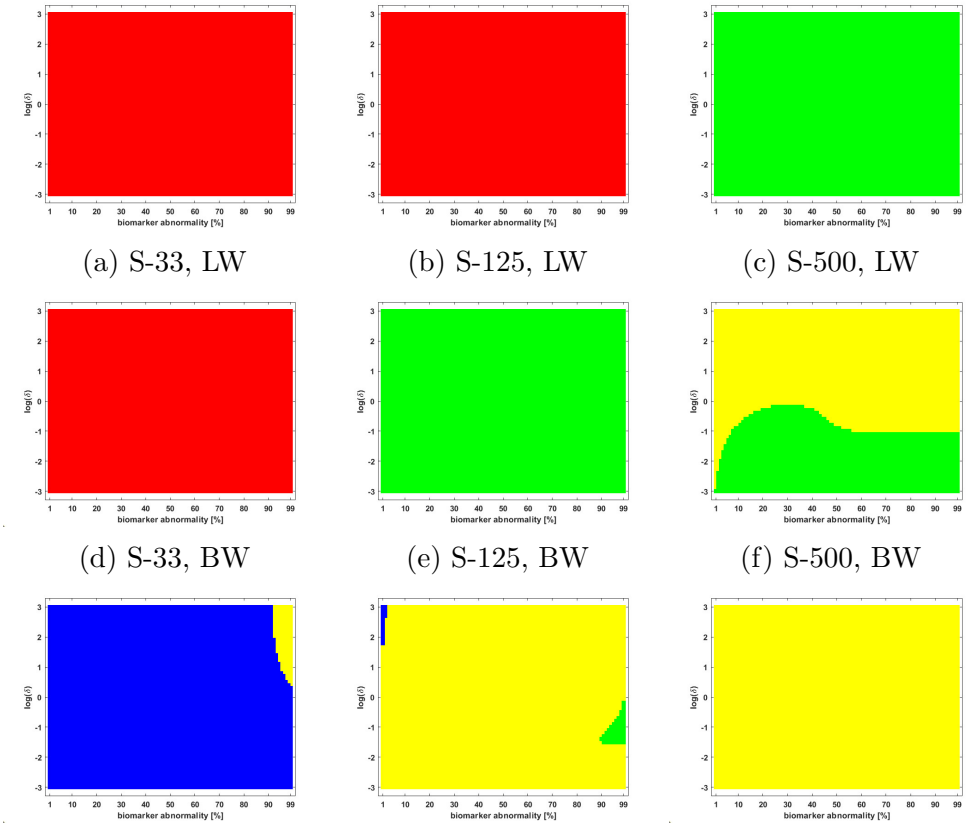
Figure 2: Braid surfaces, observed computational τP NFT staging with deterministic streamlined connectomes; diffusion dominated regime ($\ln(\beta) = 2$). Length-free (top), ballistic (middle) and diffusive (bottom) weighting schemes. The x-axis determines the biomarker abnormality threshold $1\% < T \leq 100\%$ and the y-axis corresponds to NFT aggregation rate (δ) with $-3 \leq \ln(\delta) \leq 3$.



(a) S-33, BW (b) S-125, BW (c) S-500, BW

(d) S-33, DW (e) S-125, DW (f) S-500, DW

Figure 3: Braid surfaces, observed computational τP NFT staging with deterministic streamlined connectomes; diffusion dominated regime ($\ln(\beta) = 2$). Length-free (top), ballistic (middle) and diffusive (bottom) weighting schemes. The x-axis determines the biomarker abnormality threshold $1\% < T \leq 100\%$ and the y-axis corresponds to NFT aggregation rate (δ) with $-3 \leq \ln(\delta) \leq 3$.



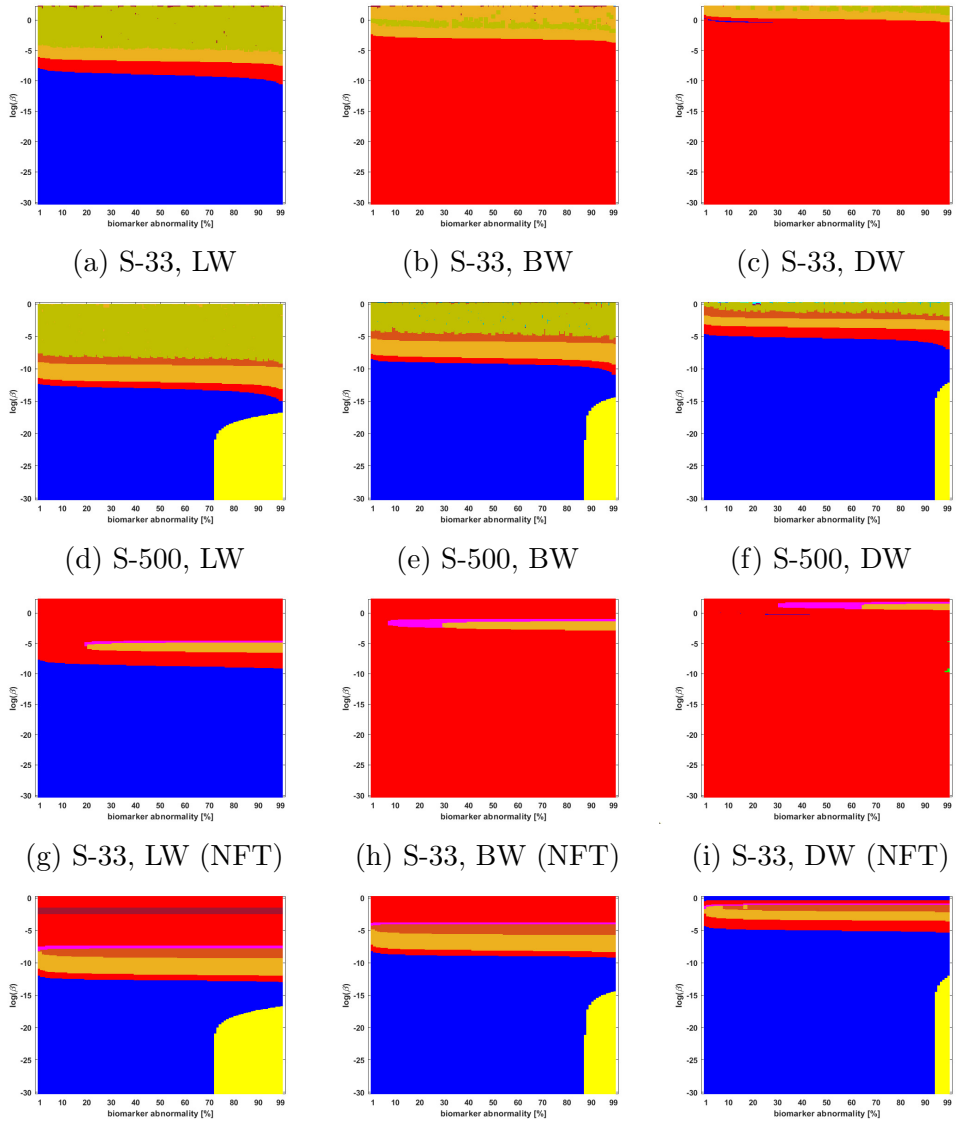
(a) S-33, LW (b) S-125, LW (c) S-500, LW

(d) S-33, BW (e) S-125, BW (f) S-500, BW

(g) S-33, DW (h) S-125, DW (i) S-500, DW

Figure 4: Braid surfaces, observed computational τP NFT staging with deterministic streamlined connectomes; growth dominated regime ($\ln(\beta) = -3$). Length-free (top), ballistic (middle) and diffusive (bottom) weighting schemes. The x-axis determines the biomarker abnormality threshold $1\% < T \leq 100\%$ and the y-axis corresponds to NFT aggregation rate (δ) with $-3 \leq \ln(\delta) \leq 3$.

Additional results, probabilistic streamlined connectome staging



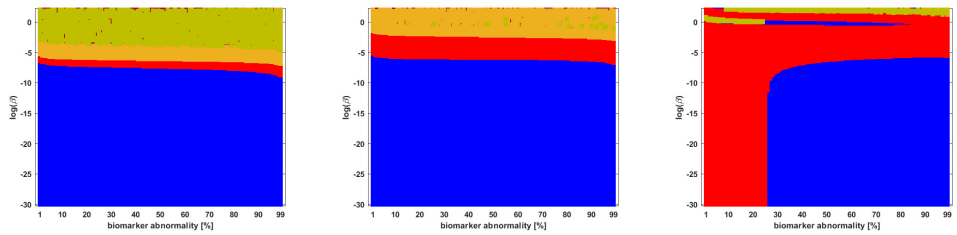
(a) S-33, LW (b) S-33, BW (c) S-33, DW

(d) S-500, LW (e) S-500, BW (f) S-500, DW

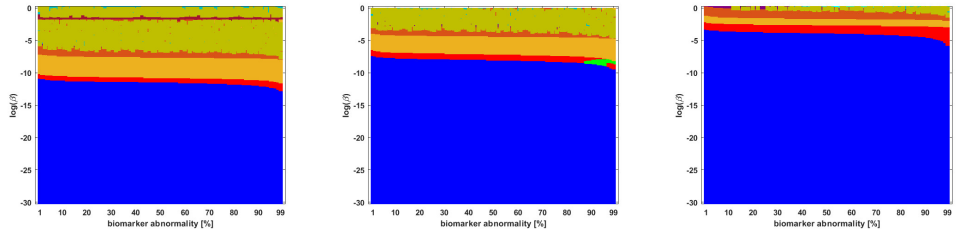
(g) S-33, LW (NFT) (h) S-33, BW (NFT) (i) S-33, DW (NFT)

(j) S-500, LW (NFT) (k) S-500, BW (NFT) (l) S-500, DW (NFT)

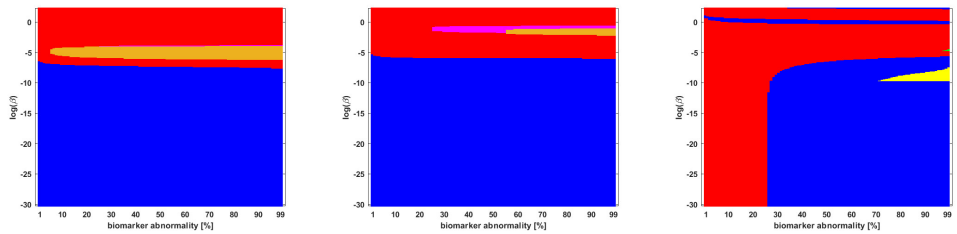
Figure 5: Observed computational (probabilistic) connectome τP seed staging (top two rows) and τP NFT staging (bottom two rows). Density filter thresholding at a threshold of 8×10^{-1} with biomarker abnormality $1\% \leq T \leq 100\%$ (x-axis) and $-30 \leq \ln(\beta) \leq 0$ (y-axis)



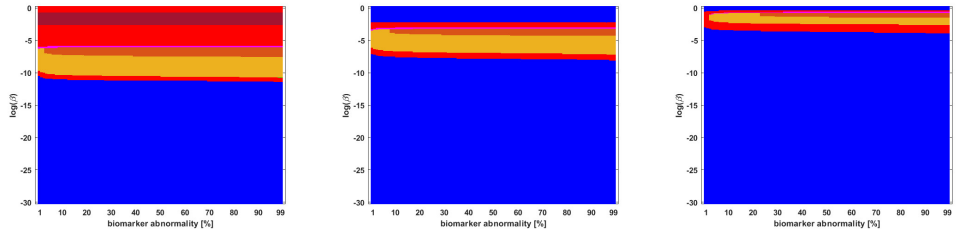
(a) S-33, LW (b) S-33, BW (c) S-33, DW



(d) S-500, LW (e) S-500, BW (f) S-500, DW

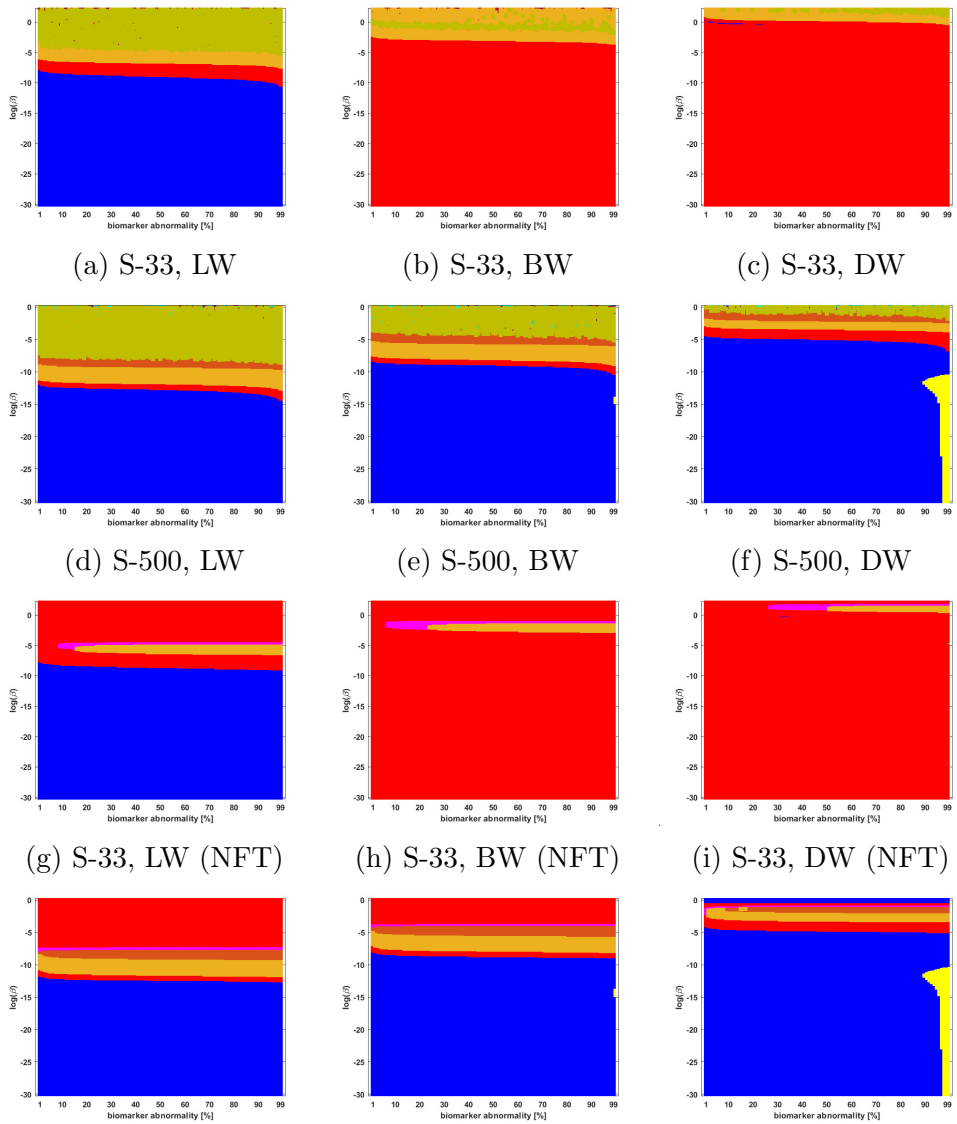


(g) S-33, LW (NFT) (h) S-33, BW (NFT) (i) S-33, DW (NFT)



(j) S-500, LW (NFT) (k) S-500, BW (NFT) (l) S-500, DW (NFT)

Figure 6: Observed computational (probabilistic) connectome τP seed staging (top two rows) and τP NFT staging (bottom two rows). High salience skeleton at a threshold of 5×10^{-4} with biomarker abnormality $1\% \leq T \leq 100\%$ (x-axis) and $-30 \leq \ln(\beta) \leq 0$ (y-axis)



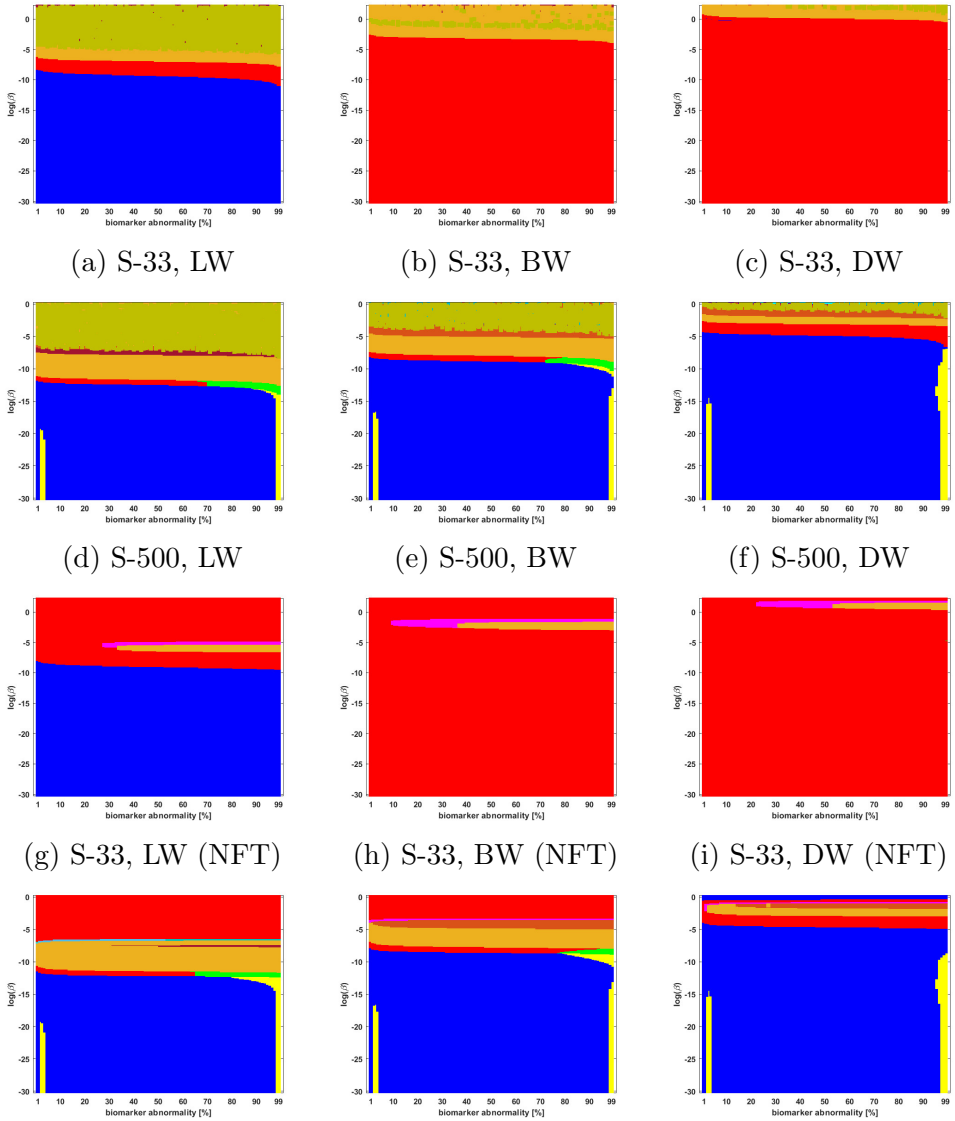
(a) S-33, LW (b) S-33, BW (c) S-33, DW

(d) S-500, LW (e) S-500, BW (f) S-500, DW

(g) S-33, LW (NFT) (h) S-33, BW (NFT) (i) S-33, DW (NFT)

(j) S-500, LW (NFT) (k) S-500, BW (NFT) (l) S-500, DW (NFT)

Figure 7: Observed computational (probabilistic) connectome τP seed staging (top two rows) and τP NFT staging (bottom two rows). Noise corrected backbone at a threshold of 1.28 with biomarker abnormality $1\% \leq T \leq 100\%$ (x-axis) and $-30 \leq \ln(\beta) \leq 0$ (y-axis)



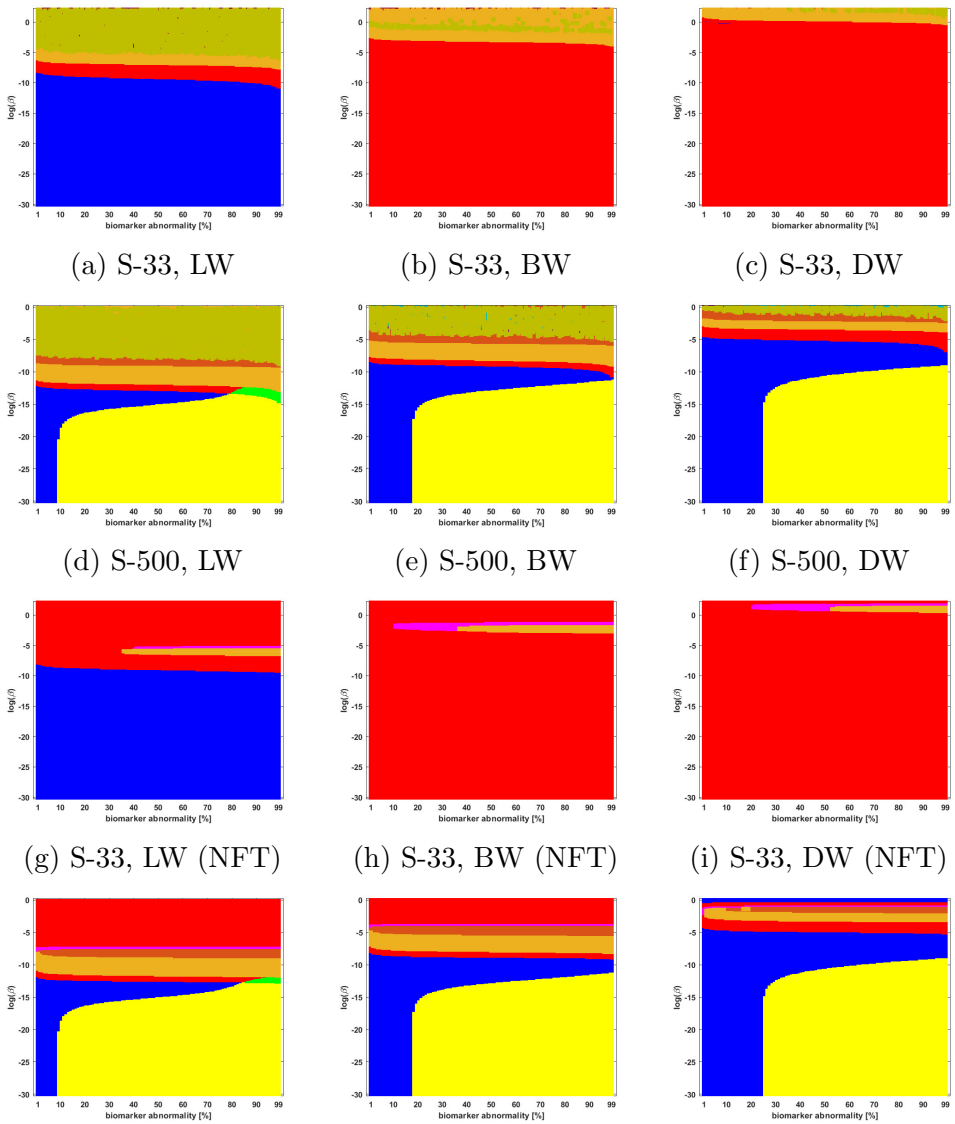
(a) S-33, LW (b) S-33, BW (c) S-33, DW

(d) S-500, LW (e) S-500, BW (f) S-500, DW

(g) S-33, LW (NFT) (h) S-33, BW (NFT) (i) S-33, DW (NFT)

(j) S-500, LW (NFT) (k) S-500, BW (NFT) (l) S-500, DW (NFT)

Figure 8: Observed computational (probabilistic) connectome τP seed staging (top two rows) and τP NFT staging (bottom two rows). Naive thresholding at a threshold of 5×10^{-3} with biomarker abnormality $1\% \leq T \leq 100\%$ (x-axis) and $-30 \leq \ln(\beta) \leq 0$ (y-axis)



(a) S-33, LW (b) S-33, BW (c) S-33, DW

(d) S-500, LW (e) S-500, BW (f) S-500, DW

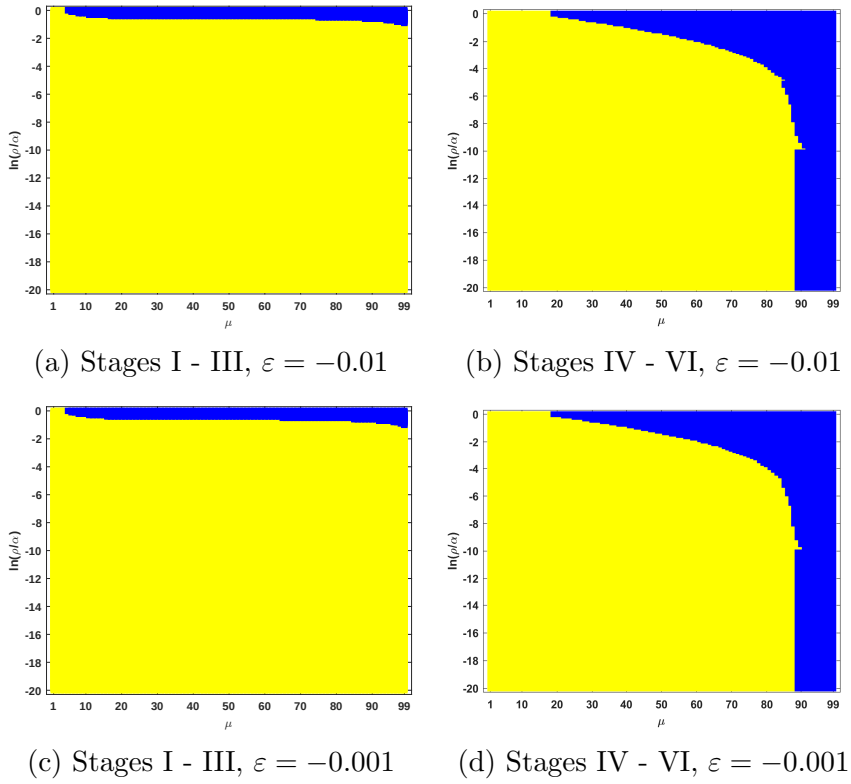
(g) S-33, LW (NFT) (h) S-33, BW (NFT) (i) S-33, DW (NFT)

(j) S-500, LW (NFT) (k) S-500, BW (NFT) (l) S-500, DW (NFT)

Figure 9: Observed computational (probabilistic) connectome τP seed staging (top two rows) and τP NFT staging ($\delta = 1$, bottom two rows). Naive thresholding at a threshold of 1×10^{-3} with biomarker abnormality $1\% \leq T \leq 100\%$ (x-axis) and $-30 \leq \ln(\beta) \leq 0$ (y-axis)

Appendix B
Additional Staging Results of Chapter 5

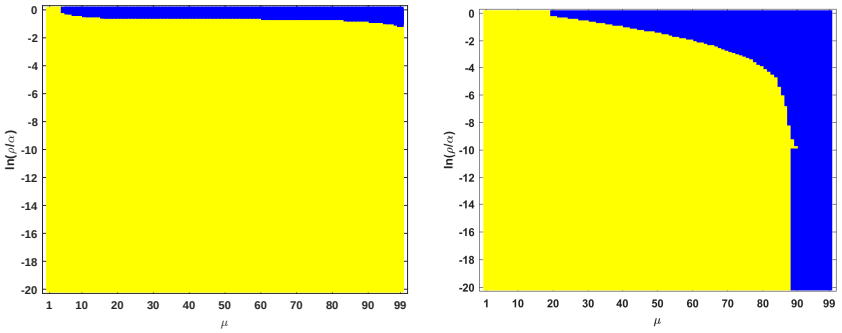
Additional results, dynamics of heterogeneous system with regional carrying capacities



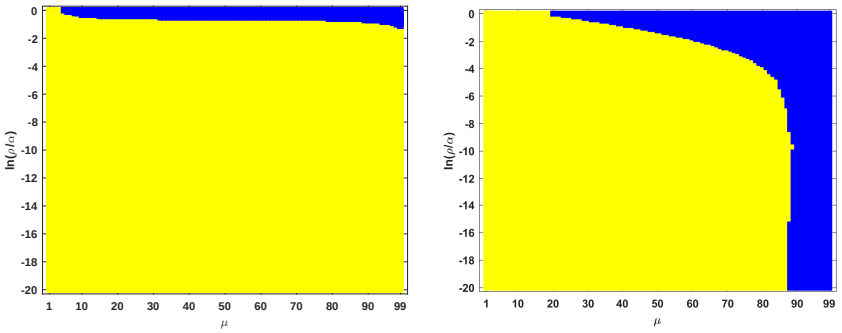
(a) Stages I - III, $\varepsilon = -0.01$ (b) Stages IV - VI, $\varepsilon = -0.01$

(c) Stages I - III, $\varepsilon = -0.001$ (d) Stages IV - VI, $\varepsilon = -0.001$

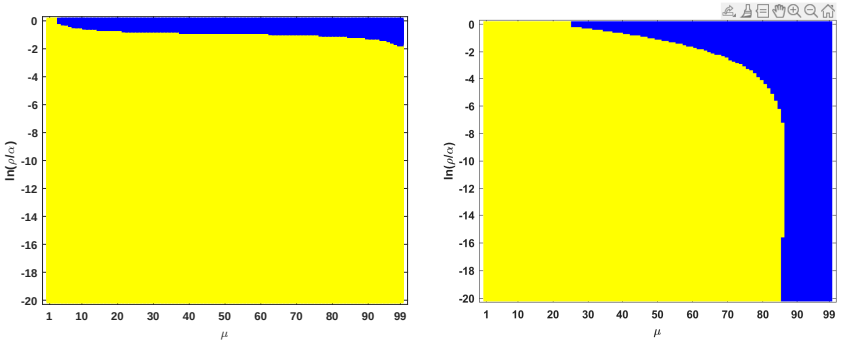
Figure 10: Observed computational (deterministic) connectome τP seed staging; Stages I - III (left) and stages IV - VI (right). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (5.33).



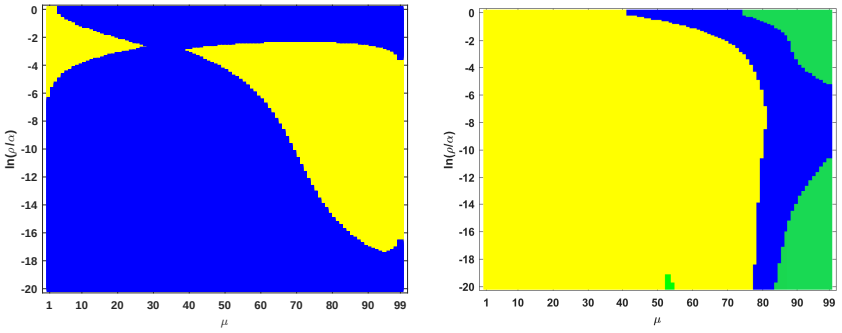
(a) Stages I - III, $\varepsilon = 0.001$ (b) Stages IV - VI, $\varepsilon = 0.001$



(c) Stages I - III, $\varepsilon = 0.01$ (d) Stages IV - VI, $\varepsilon = 0.01$



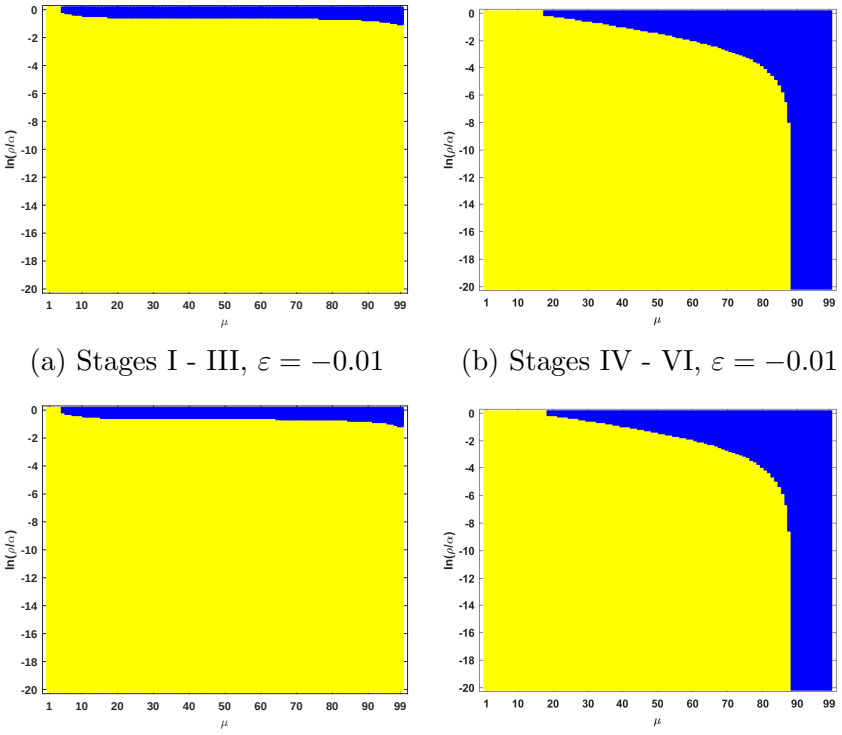
(e) Stages I - III, $\varepsilon = 0.1$ (f) Stages IV - VI, $\varepsilon = 0.1$



(g) Stages I - III, $\varepsilon = 0.5$ (h) Stages IV - VI, $\varepsilon = 0.5$

Figure 11: Observed computational (deterministic) connectome τP seed staging; Stages I - III (left) and stages IV - VI (right). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (5.33).

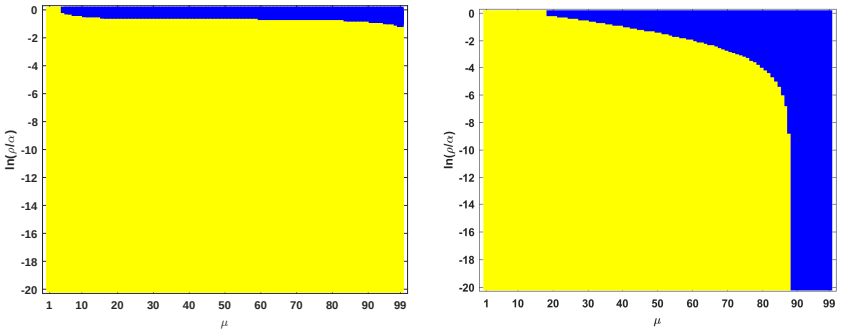
Additional results, dynamics of heterogeneous system with regional growth rates



(a) Stages I - III, $\varepsilon = -0.01$ (b) Stages IV - VI, $\varepsilon = -0.01$

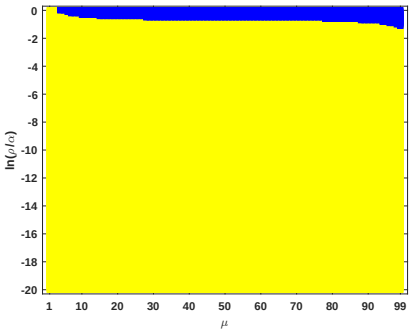
(c) Stages I - III, $\varepsilon = -0.001$ (d) Stages IV - VI, $\varepsilon = -0.001$

Figure 12: Observed computational (deterministic) connectome τP seed staging; Stages I - III (left) and stages IV - VI (right). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (5.34).

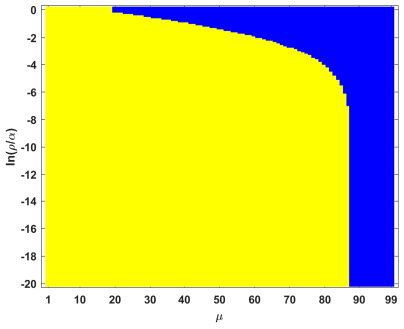


(a) Stages I - III, $\varepsilon = 0.001$

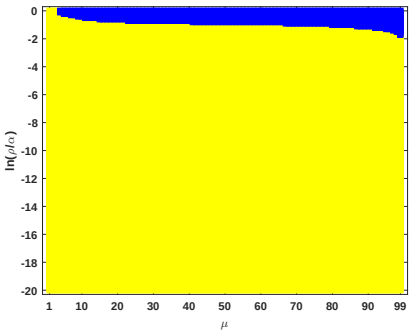
(b) Stages IV - VI, $\varepsilon = 0.001$



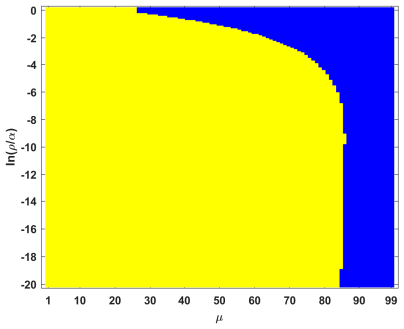
(c) Stages I - III, $\varepsilon = 0.01$



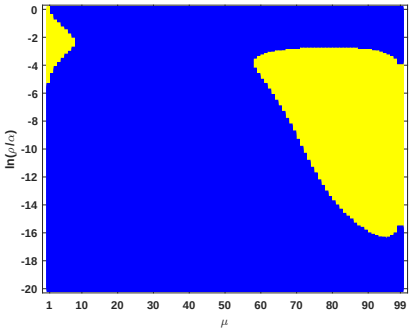
(d) Stages IV - VI, $\varepsilon = 0.01$



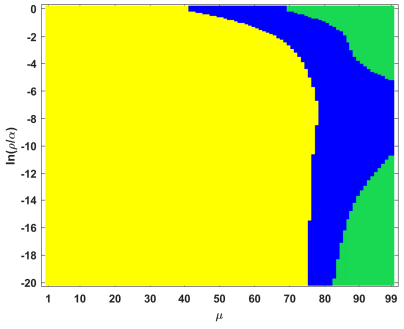
(e) Stages I - III, $\varepsilon = 0.1$



(f) Stages IV - VI, $\varepsilon = 0.1$



(g) Stages I - III, $\varepsilon = 0.5$

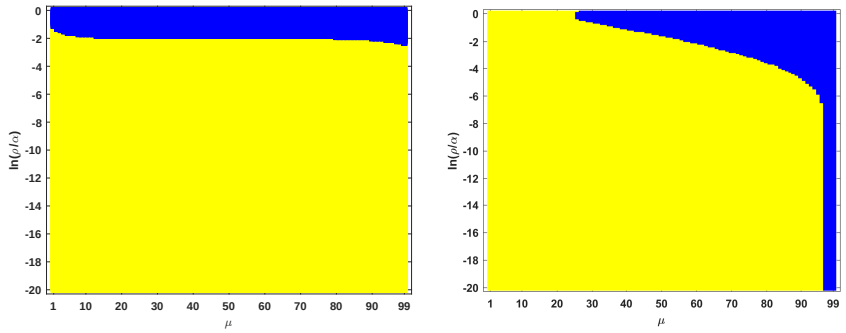


(h) Stages IV - VI, $\varepsilon = 0.5$

Figure 13: Observed computational (deterministic) connectome τP seed staging; Stages I - III (left) and stages IV - VI (right). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (5.34).

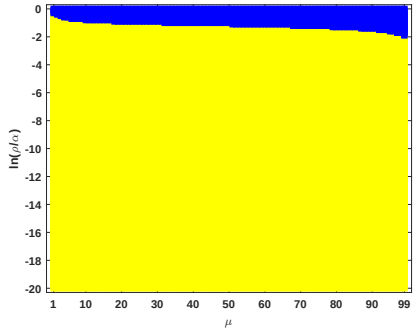
Appendix C
Additional Staging Results of Chapter 6

Additional results, dynamics on directed networks from node degree multiplication

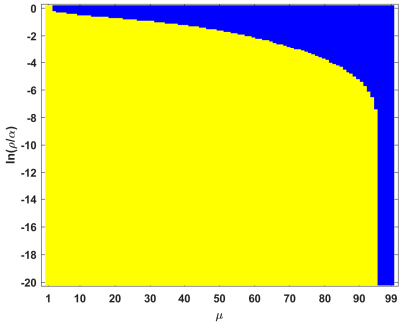


(a) Stages I - III, $\varepsilon = -0.5$

(b) Stages IV - VI, $\varepsilon = -0.5$

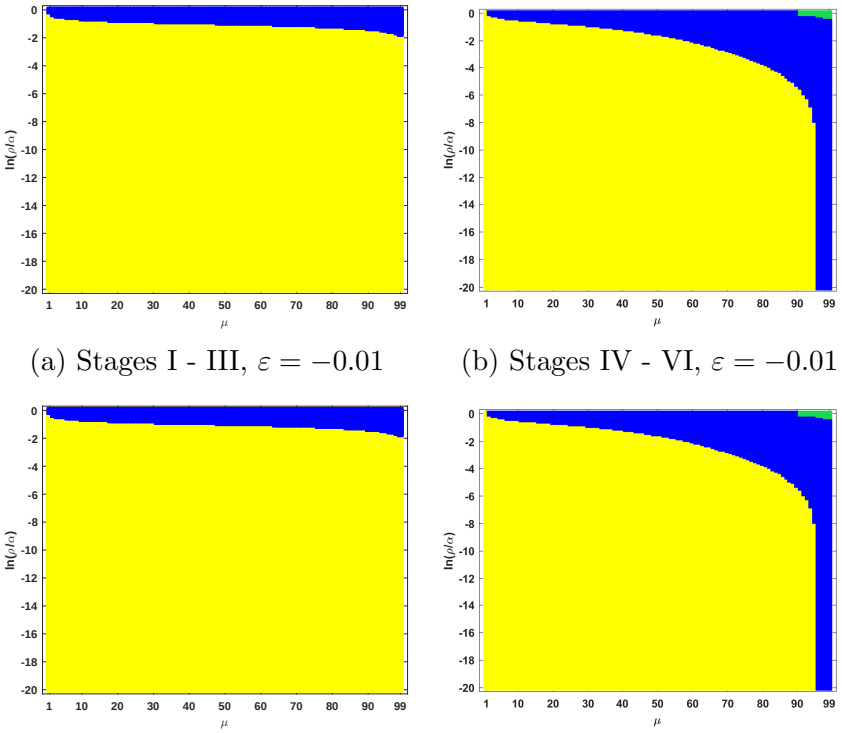


(c) Stages I - III, $\varepsilon = -0.1$



(d) Stages IV - VI, $\varepsilon = -0.1$

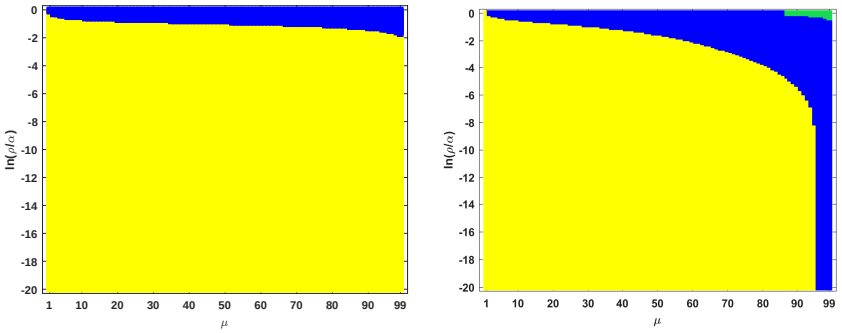
Figure 14: Observed computational (deterministic) connectome τP seed staging; Stages I - III (left) and stages IV - VI (right). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (6.12) with $\mathbf{L}^{\text{dir}}$(6.7) from $\mathbf{A}^{\text{dir}}$ (6.37).



(a) Stages I - III, $\varepsilon = -0.01$ (b) Stages IV - VI, $\varepsilon = -0.01$

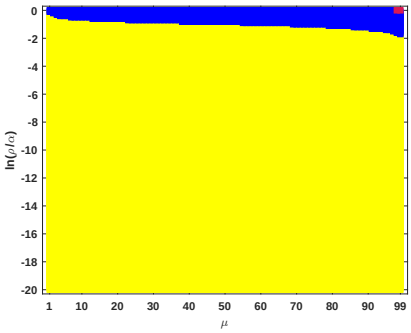
(c) Stages I - III, $\varepsilon = -0.001$ (d) Stages IV - VI, $\varepsilon = -0.001$

Figure 15: Observed computational (deterministic) connectome τP seed staging; Stages I - III (left) and stages IV - VI (right). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (6.12) with $\mathbf{L}^{\text{dir}}$(6.7) from $\mathbf{A}^{\text{dir}}$ (6.37).

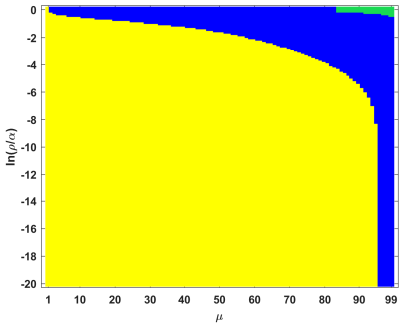


(a) Stages I - III, $\varepsilon = 0.001$

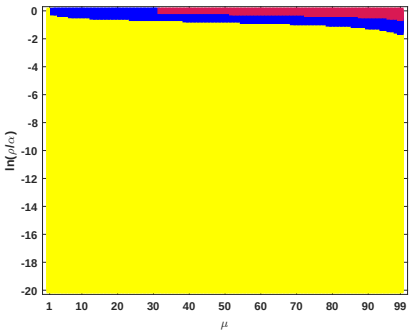
(b) Stages IV - VI, $\varepsilon = 0.001$



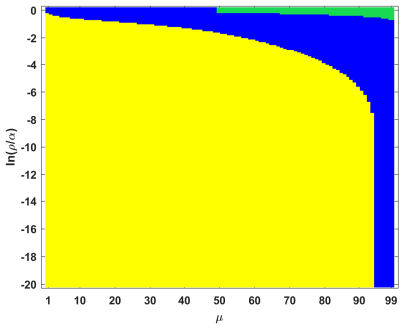
(c) Stages I - III, $\varepsilon = 0.01$



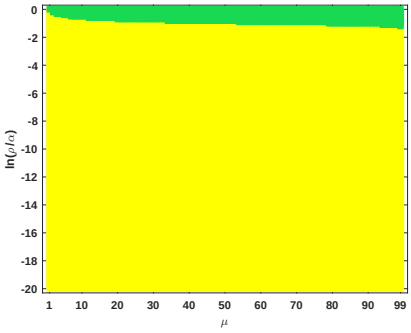
(d) Stages IV - VI, $\varepsilon = 0.01$



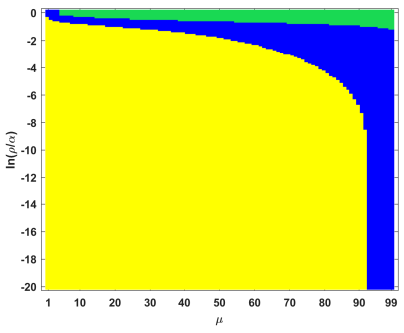
(e) Stages I - III, $\varepsilon = 0.1$



(f) Stages IV - VI, $\varepsilon = 0.1$



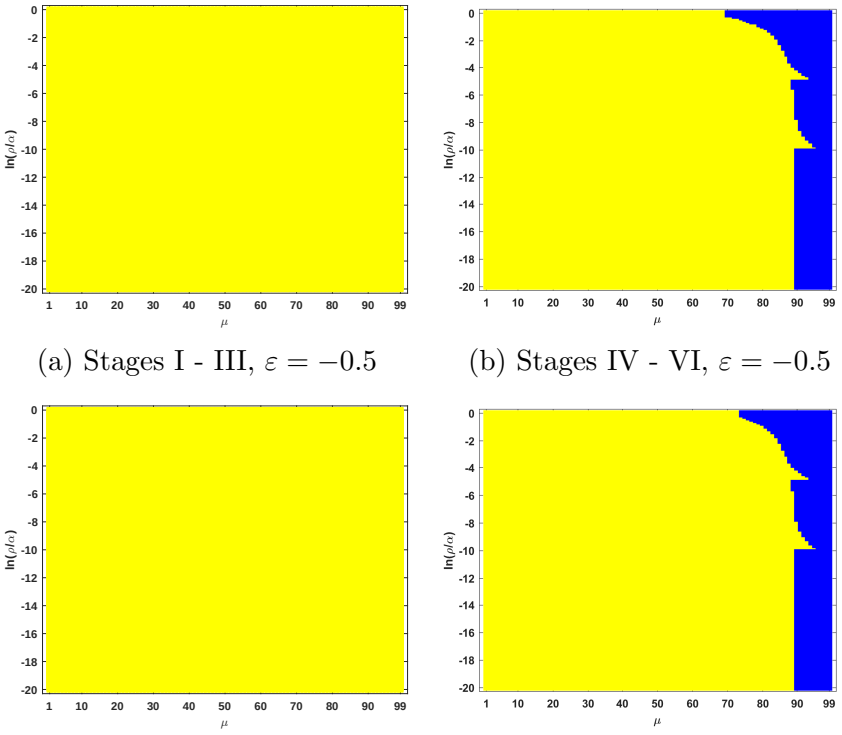
(g) Stages I - III, $\varepsilon = 0.5$



(h) Stages IV - VI, $\varepsilon = 0.5$

Figure 16: Observed computational (deterministic) connectome τP seed staging; Stages I - III (left) and stages IV - VI (right). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (6.12) with $\mathbf{L}^{\text{dir}}$(6.7) from $\mathbf{A}^{\text{dir}}$ (6.37).

Additional results, dynamics on directed networks from carrying capacity multiplication



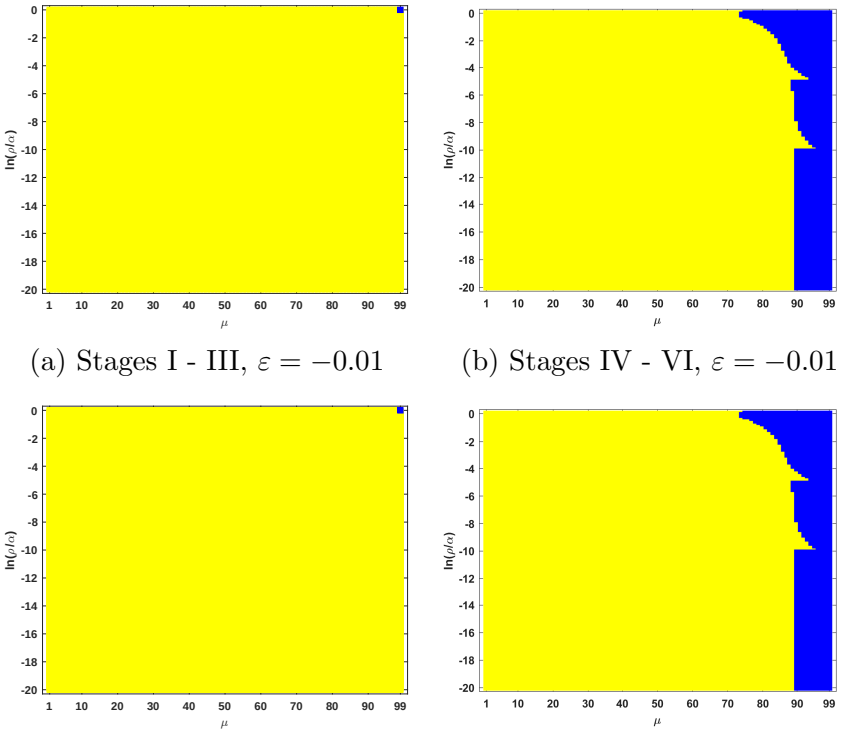
(a) Stages I - III, $\varepsilon = -0.5$

(b) Stages IV - VI, $\varepsilon = -0.5$

(c) Stages I - III, $\varepsilon = -0.1$

(d) Stages IV - VI, $\varepsilon = -0.1$

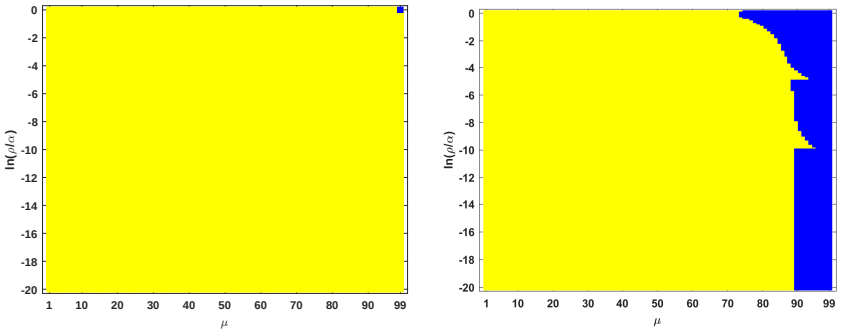
Figure 17: Observed computational (deterministic) connectome τP seed staging; Stages I - III (left) and stages IV - VI (right). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (6.12) with $\mathbf{L}^{\text{dir}}$(6.7) from $\mathbf{A}^{\text{dir}}$ (6.38).



(a) Stages I - III, $\varepsilon = -0.01$ (b) Stages IV - VI, $\varepsilon = -0.01$

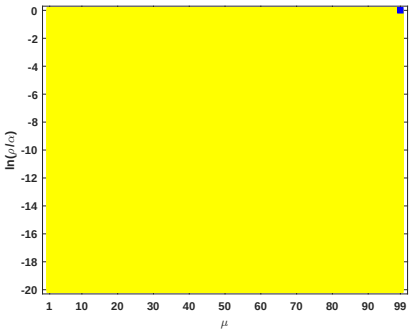
(c) Stages I - III, $\varepsilon = -0.001$ (d) Stages IV - VI, $\varepsilon = -0.001$

Figure 18: Observed computational (deterministic) connectome τP seed staging; Stages I - III (left) and stages IV - VI (right). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (6.12) with $\mathbf{L}^{\text{dir}}$(6.7) from $\mathbf{A}^{\text{dir}}$ (6.38).

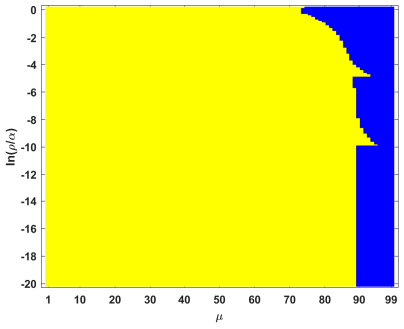


(a) Stages I - III, $\varepsilon = 0.001$

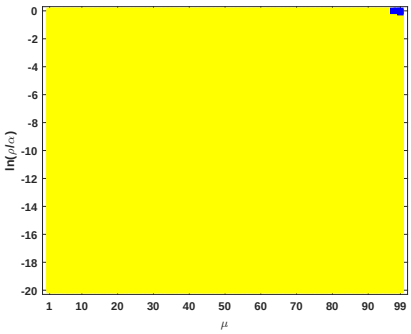
(b) Stages IV - VI, $\varepsilon = 0.001$



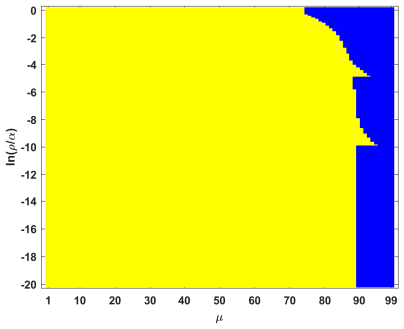
(c) Stages I - III, $\varepsilon = 0.01$



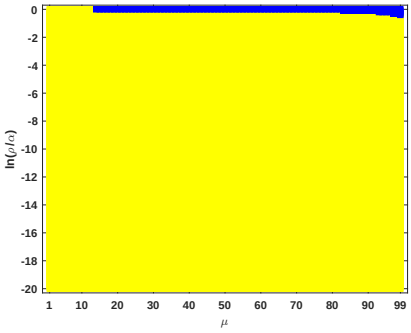
(d) Stages IV - VI, $\varepsilon = 0.01$



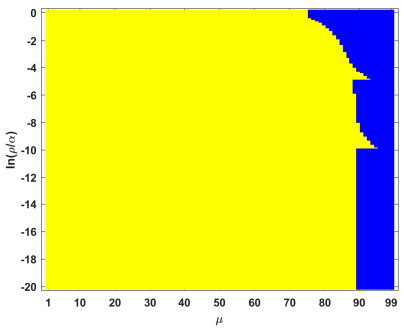
(e) Stages I - III, $\varepsilon = 0.1$



(f) Stages IV - VI, $\varepsilon = 0.1$



(g) Stages I - III, $\varepsilon = 0.5$



(h) Stages IV - VI, $\varepsilon = 0.5$

Figure 19: Observed computational (deterministic) connectome τP seed staging; Stages I - III (left) and stages IV - VI (right). The x-axis determines the biomarker abnormality threshold $1\% \leq \mu < 100\%$ and the y-axis corresponds to $-20 \leq \ln(\rho/\alpha) \leq 0$ for the parameter ρ and α in (6.12) with $\mathbf{L}^{\text{dir}}$(6.7) from $\mathbf{A}^{\text{dir}}$ (6.38).

Bibliography

[1] Farras Abdelnour, Henning U Voss, and Ashish Raj. Network diffusion accurately models the relationship between structural and functional brain connectivity networks. *Neuroimage*, 90:335–347, 2014.

[2] Flortaucipir (AV-1451) processing methods. `http://adni.loni.usc.edu/data-samples/access-data/`. Accessed: 20-12-2021.

[3] Mutations. `https://www.alzforum.org/mutations`. Accessed: 2022-03-24.

[4] Athena Andreadis. Misregulation of tau alternative splicing in neurodegeneration and dementia. *Alternative splicing and disease*, pages 89–107, 2006.

[5] Ashley Armbruster, Matt Holzer, Noah Roselli, and Lena Underwood. Epidemic spreading on complex networks as front propagation into an unstable state. *Bulletin of Mathematical Biology*, 85(1):4, 2023.

[6] Karen H Ashe and Adriano Aguzzi. Prions, prionoids and pathogenic proteins in Alzheimer disease. *Prion*, 7(1):55–59, 2013.

[7] Selen Atasoy, Isaac Donnelly, and Joel Pearson. Human brain networks function in connectome-specific harmonic waves. *Nature Communications*, 7:10340, 2016.

[8] Selen Atasoy, Leor Roseman, Mendel Kaelen, Morten L Kringelbach, Gustavo Deco, and Robin L Carhart-Harris. Connectome-harmonic decomposition of human brain activity reveals dynamical repertoire re-organization under lsd. *Scientific Reports*, 7(1):17661, 2017.

[9] S.L. Baker et al. Reference Tissue-Based Kinetic Evaluation of 18F-AV1451 for Tau Imaging. *J. Nucl. Med.*, 58(2):332–338, 2017.

[10] S.L. Baker, A. Maass, and W.J. Jagust. Considerations and code for partial volume correcting [(18F)]-AV-1451 tau PET data. *Data Brief*, 15:648–657, 2017.

[11] Farah H Bardai, Liqun Wang, Yamini Mutreja, Mythili Yenjerla, T Chris Gamblin, and Mel B Feany. A conserved cytoskeletal signaling cascade mediates neurotoxicity of ftdp-17 tau mutations in vivo. *Journal of Neuroscience*, 38(1):108–119, 2018.

[12] Timothy EJ Behrens, H Johansen Berg, Saad Jbabdi, Matthew FS Rushworth, and Mark W Woolrich. Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *neuroimage*, 34(1):144–155, 2007.

[13] Alexandre Bejanin, Daniel R Schonhaut, Renaud La Joie, Joel H Kramer, Suzanne L Baker, Natasha Sosa, Nagehan Ayakta, Averill Cantwell, Mustafa Janabi, Mariella Lauriola, et al. Tau pathology and neurodegeneration contribute to cognitive impairment in alzheimer's disease. *Brain*, 140(12):3286–3300, 2017.

[14] Christophe Besse and Grégory Faye. Spreading properties for sir models on homogeneous trees. *Bulletin of Mathematical Biology*, 83(11):1–27, 2021.

[15] Kaj Blennow, Henrik Zetterberg, and Anne M Fagan. Fluid biomarkers in Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, 2(9):a006221, 2012.

[16] Elin S Blom, Vilmantas Giedraitis, Henrik Zetterberg, Hiroaki Fukumoto, Kaj Blennow, Bradley T Hyman, Michael C Irizarry, Lars-Olof Wahlund, Lars Lannfelt, and Martin Ingelsson. Rapid progression from mild cognitive impairment to alzheimer's disease in subjects with elevated levels of tau in cerebrospinal fluid and the apoe $\varepsilon 4/\varepsilon 4$ genotype. *Dementia and geriatric cognitive disorders*, 27(5):458–464, 2009.

[17] Heiko Braak, Irina Alafuzoff, Thomas Arzberger, Hans Kretzschmar, and Kelly Del Tredici. Staging of Alzhimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathologica*, 112(4):389–404, 2006.

[18] Heiko Braak and Eva Braak. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, 82(4):239–259, 1991.

[19] Heiko Braak and Kelly Del Tredici. *Neuroanatomy and pathology of sporadic Alzheimer's disease*. Springer, 2014.

[20] Heiko Braak, Kelly Del Tredici, Udo Rüb, Rob AI De Vos, Ernst NH Jansen Steur, and Eva Braak. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiology of Aging*, 24(2):197–211, 2003.

[21] Paul C Bressloff and Jay M Newby. Stochastic models of intracellular transport. *Reviews of Modern Physics*, 85(1):135, 2013.

[22] Jessica Bright, Sami Hussain, Vu Dang, Sarah Wright, Bonnie Cooper, Tony Byun, Carla Ramos, Andrew Singh, Graham Parry, Nancy Stagliano, et al. Human secreted tau increases amyloid-beta production. *Neurobiology of aging*, 36(2):693–709, 2015.

[23] Dirk Brockmann and Dirk Helbing. The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342(6164):1337–1342, 2013.

[24] Kathryn J Bryan, Hyoung-gon Lee, George Perry, Mark A Smith, and Gemma Casadesus. Transgenic mouse models of alzheimer's disease: behavioral testing and considerations. chapter 1. Crc Press, 2011.

[25] Rachel F Buckley, Bernard Hanseeuw, Aaron P Schultz, Patrizia Vannini, Sarah L Aghjayan, Michael J Properzi, Jonathan D Jackson, Elizabeth C Mormino, Dorene M Rentz, Reisa A Sperling, et al. Region-specific association of subjective cognitive decline with tauopathy independent of global β-amyloid burden. *JAMA neurology*, 74(12):1455–1463, 2017.

[26] Rachel F Buckley, Paul Maruff, David Ames, Pierrick Bourgeat, Ralph N Martins, Colin L Masters, Stephanie Rainey-Smith, Nicola Lautenschlager, Christopher C Rowe, Greg Savage, et al. Subjective memory decline predicts greater rates of clinical progression in preclinical alzheimer's disease. *Alzheimer's & Dementia*, 12(7):796–804, 2016.

[27] Randy L Buckner, Abraham Z Snyder, Benjamin J Shannon, Gina LaRossa, Rimmon Sachs, Anthony F Fotenos, Yvette I Sheline, William E Klunk, Chester A Mathis, John C Morris, et al. Molecular, structural, and functional characterization of Alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. *Journal of Neuroscience*, 25(34):7709–7717, 2005.

[28] Orso Bugiani, Giorgio Giaccone, Blas Frangione, Bernardino Ghetti, and Fabrizio Tagliavini. Alzheimer patients: preamyloid deposits are more widely distributed than senile plaques throughout the central nervous system. *Neuroscience letters*, 103(3):263–268, 1989.

[29] Keith Burghardt, Siyi Guo, and Kristina Lerman. Unequal impact and spatial aggregation distort covid-19 growth rates. *Philosophical Transactions of the Royal Society A*, 380(2214):20210122, 2022.

[30] 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.

[31] 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 Hyoung Lyoo. In vivo cortical spreading pattern of tau and amyloid in the Alzheimer disease spectrum. *Annals of Neurology*, 80(2):247–258, 2016.

[32] Jae Yong Choi, Hanna Cho, Sung Jun Ahn, Jae Hoon Lee, Young Hoon Ryu, Myung Sik Lee, and Chul Hyoung Lyoo. Off-target 18f-av-1451 binding in the basal ganglia correlates with age-related iron accumulation. *Journal of Nuclear Medicine*, 59(1):117–120, 2018.

[33] Fan RK Chung and Fan Chung Graham. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.

[34] Florence Clavaguera, Tristan Bolmont, R Anthony Crowther, Dorothee Abramowski, Stephan Frank, Alphonse Probst, Graham Fraser, Anna K Stalder, Martin Beibel, Matthias Staufenbiel, et al. Transmission and spreading of tauopathy in transgenic mouse brain. *Nature cell biology*, 11(7):909–913, 2009.

[35] Florence Clavaguera, Isabelle Lavenir, Ben Falcon, Stephan Frank, Michel Goedert, and Markus Tolnay. "prion-like" templated misfolding in tauopathies. *Brain Pathology*, 23(3):342–349, 2013.

[36] Jonathan E Conklin, John V Lieberman, Cathleen A Barnes, and Daniel Z Louis. Disease staging: implications for hospital reimbursement and management. *Health care financing review*, 1984(Suppl):13, 1984.

[37] Eli J Cornblath, Howard L Li, Lakshmi Changolkar, Bin Zhang, Hannah J Brown, Ronald J Gathagan, Modupe F Olufemi, John Q Trojanowski, Danielle S Bassett, Virginia MY Lee, et al. Computational modeling of tau pathology spread reveals patterns of regional vulnerability and the impact of a genetic risk factor. *Science Advances*, 7(24):eabg6677, 2021.

[38] Michele Coscia and Frank Neffke. Network backboning with noisy data. In *International Conference on Data Engineering (ICDE)*, 2017.

[39] Paolo d 'Errico and Melanie Meyer-Luehmann. Mechanisms of pathogenic tau and aβ protein spreading in alzheimer's disease. *Frontiers in Aging Neuroscience*, 12:265, 2020.

[40] A. Daducci, S. Gerhard, and J.-P. et. al. Thiran. The Connectome Mapper: An Open-Source Processing Pipeline to Map Connectomes with MRI. *PLoS One*, 7(12):e48121, 2012.

[41] André Delacourte, Jean-Philippe David, Nicolas Sergeant, L Buee, A Wattez, P Vermersch, 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.

[42] J-C Delvenne, Sophia N Yaliraki, and Mauricio Barahona. Stability of graph communities across time scales. *Proceedings of the national academy of sciences*, 107(29):12755–12760, 2010.

[43] Michael A DeTure and Dennis W Dickson. The neuropathological diagnosis of alzheimer's disease. *Molecular neurodegeneration*, 14(1):1–18, 2019.

[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] An-Tao Du, Norbert Schuff, Joel H Kramer, Howard J Rosen, Maria Luisa Gorno-Tempini, Katherine Rankin, Bruce L Miller, and Michael W Weiner. Different regional patterns of cortical thinning in alzheimer's disease and frontotemporal dementia. *Brain*, 130(4):1159–1166, 2007.

[46] P Edison, HA Archer, R Hinz, A Hammers, N Pavese, YF Tai, G Hotton, D Cutler, N Fox, A Kennedy, et al. Amyloid, hypometabolism, and cognition in alzheimer disease: an [11c] pib and [18f] fdg pet study. *Neurology*, 68(7):501–508, 2007.

[47] Manfred Eigen. Prionics or the kinetic basis of prion diseases. *Biophysical chemistry*, 63(1):A1–A18, 1996.

[48] Grégory Faye. Traveling fronts for lattice neural field equations. *Physica D: Nonlinear Phenomena*, 378-379:20–32, 2018.

[49] 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.

[50] 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.

[51] 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.

[52] François Fouss, Marco Saerens, and Masashi Shimbo. *Algorithms and models for network data and link analysis*. Cambridge University Press, 2016.

[53] Ralf P Friedrich, Katharina Tepper, Raik Rönicke, Malle Soom, Martin Westermann, Klaus Reymann, Christoph Kaether, and Marcus Fändrich. Mechanism of amyloid plaque formation suggests an intracellular basis of aβ pathogenicity. *Proceedings of the National Academy of Sciences*, 107(5):1942–1947, 2010.

[54] Bess Frost, Rachel L Jacks, and Marc I Diamond. Propagation of tau misfolding from the outside to the inside of a cell. *Journal of Biological Chemistry*, 284(19):12845–12852, 2009.

[55] Hongjun Fu, John Hardy, and Karen E Duff. Selective vulnerability in neurodegenerative diseases. *Nature neuroscience*, 21(10):1350–1358, 2018.

[56] Hongjun Fu, Andrea Possenti, Rosie Freer, Yoshikazu Nakano, Nancy C Hernandez Villegas, Maoping Tang, Paula VM Cauhy, Benjamin A Lassus, Shuo Chen, Stephanie L Fowler, et al. A tau homeostasis signature is linked with the cellular and regional vulnerability of excitatory neurons to tau pathology. *Nature neuroscience*, 22(1):47–56, 2019.

[57] 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.

[58] Conditional inference trees in r programming. `https://www.geeksforgeeks.org/conditional-inference-trees-in-r-programming/`. Accessed: 13-02-2023.

[59] MGSM Goedert, MG Spillantini, R Jakes, D Rutherford, and RA Crowther. Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of alzheimer's disease. *Neuron*, 3(4):519–526, 1989.

[60] Michel Goedert. Alzheimer's and Parkinson's diseases: The prion concept in relation to assembled aβ, tau, and α-synuclein. *Science*, 349(6248):1255555, 2015.

[61] Teresa Gómez-Isla, Richard Hollister, Howard West, Stina Mui, John H Growdon, Ronald C Petersen, Joseph E Parisi, and Bradley T Hyman. Neuronal loss correlates with but exceeds neurofibrillary tangles in alzheimer's disease. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 41(1):17–24, 1997.

[62] Joseph S Gonnella, Mark C Hornbrook, and Daniel Z Louis. Staging of disease: a case-mix measurement. *Jama*, 251(5):637–644, 1984.

[63] Jürgen Götz, Glenda Halliday, and Rebecca M Nisbet. Molecular pathogenesis of the tauopathies. *Annual Review of Pathology: Mechanisms of Disease*, 14:239–261, 2019.

[64] D. Grady, C. Thiemann, and D. Brockmann. Robust classification of salient links in complex networks. *Nat. Commun.*, 3(864), 2012.

[65] Timo Grimmer, Matthias Riemenschneider, Hans Förstl, Gjermund Henriksen, William E Klunk, Chester A Mathis, Tohru Shiga, Hans-Jürgen Wester, Alexander Kurz, and Alexander Drzezga. Beta amyloid in alzheimer's disease: increased deposition in brain is reflected in reduced concentration in cerebrospinal fluid. *Biological psychiatry*, 65(11):927–934, 2009.

[66] Lei Gu and Zhefeng Guo. Alzheimer's aβ42 and aβ40 peptides form interlaced amyloid fibrils. *Journal of neurochemistry*, 126(3):305–311, 2013.

[67] Tong Guo, Wendy Noble, and Diane P Hanger. Roles of tau protein in health and disease. *Acta neuropathologica*, 133(5):665–704, 2017.

[68] Garth F Hall and Brian A Patuto. Is tau ready for admission to the prion club? *Prion*, 6(3):223–233, 2012.

[69] Martin Hallbeck, Sangeeta Nath, and Jan Marcusson. Neuron-to-neuron transmission of neurodegenerative pathology. *The Neuroscientist*, 19(6):560–566, 2013.

[70] John Hardy and David Allsop. Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends in Pharmacological Sciences*, 12:383–388, 1991.

[71] John A Hardy and Gerald A Higgins. Alzheimer's disease: the amyloid cascade hypothesis. *Science*, 256(5054):184–186, 1992.

[72] Masato Hasegawa, Michael J Smith, and Michel Goedert. Tau proteins with ftdp-17 mutations have a reduced ability to promote microtubule assembly. *FEBS letters*, 437(3):207–210, 1998.

[73] Dimitri Hefter and Andreas Draguhn. App as a protective factor in acute neuronal insults. *Frontiers in molecular neuroscience*, 10:22, 2017.

[74] Camilla Hesse, Lars Rosengren, Niels Andreasen, Pia Davidsson, Hugo Vanderstichele, Eugeen Vanmechelen, and Kaj Blennow. Transient increase in total tau but not phospho-tau in human cerebrospinal fluid after acute stroke. *Neuroscience letters*, 297(3):187–190, 2001.

[75] Aaron Hoffman and Matt Holzer. Invasion fronts on graphs: the fisher-kpp equation on homogeneous trees and erd\h {o} sr\'eyni graphs. *arXiv preprint arXiv:1610.06877*, 2016.

[76] Peter Hortschansky, Volker Schroeckh, Tony Christopeit, Giorgia Zandomeneghi, and Marcus Fändrich. The aggregation kinetics of alzheimer's β-amyloid peptide is controlled by stochastic nucleation. *Protein science*, 14(7):1753–1759, 2005.

[77] Torsten Hothorn, Kurt Hornik, and Achim Zeileis. ctree: Conditional inference trees. *The comprehensive R archive network*, 8, 2015.

[78] William T Hu, Kelly D Watts, Prashant Tailor, Trung P Nguyen, Jennifer C Howell, Raven C Lee, Nicholas T Seyfried, Marla Gearing, Chadwick M Hales, Allan I Levey, et al. Csf complement 3 and factor h are staging biomarkers in alzheimer's disease. *Acta neuropathologica communications*, 4(1):1–10, 2016.

[79] Mike Hutton, Corinne L Lendon, Patrizia Rizzu, Matt Baker, Susanne Froelich, Henry Houlden, Stuart Pickering-Brown, Sumi Chakraverty, Adrian Isaacs, Andrew Grover, et al. Association of missense and 5-splice-site mutations in tau with the inherited dementia ftdp-17. *Nature*, 393(6686):702–705, 1998.

[80] 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.

[81] Michiyo Iba, Jing L Guo, Jennifer D McBride, Bin Zhang, John Q Trojanowski, and Virginia M-Y Lee. Synthetic tau fibrils mediate transmission of neurofibrillary tangles in a transgenic mouse model of alzheimer's-like tauopathy. *Journal of Neuroscience*, 33(3):1024–1037, 2013.

[82] Milos D Ikonomovic, William E Klunk, Eric E Abrahamson, Chester A Mathis, Julie C Price, Nicholas D Tsopelas, Brian J Lopresti, Scott Ziolko, Wenzhu Bi, William R Paljug, et al. Post-mortem correlates of in vivo pib-pet amyloid imaging in a typical case of alzheimer's disease. *Brain*, 131(6):1630–1645, 2008.

[83] Madhura Ingalhalikar, Alex Smith, Drew Parker, Theodore D Satterthwaite, Mark A Elliott, Kosha Ruparel, Hakon Hakonarson, Raquel E Gur, Ruben C Gur, and Ragini Verma. Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences*, 111(2):823–828, 2014.

[84] Yasser Iturria-Medina, Félix M Carbonell, Roberto C Sotero, Francois Chouinard-Decorte, Alan C Evans, Alzheimer's Disease Neuroimaging Initiative, et al. Multifactorial causal model of brain (dis) organization and therapeutic intervention: Application to Alzheimer's disease. *Neuroimage*, 152:60–77, 2017.

[85] Yasser Iturria-Medina, Roberto C Sotero, Paule J Toussaint, Alan C Evans, and Alzheimer's Disease Neuroimaging Initiative. Epidemic spreading model to characterize misfolded proteins propagation in aging and associated neurodegenerative disorders. *PLoS Computational Biology*, 10(11):e1003956, 2014.

[86] Clifford R Jack, Terry M Therneau, Stephen D Weigand, Heather J Wiste, David S Knopman, Prashanthi Vemuri, Val J Lowe, Michelle M Mielke, Rosebud O Roberts, Mary M Machulda, et al. Prevalence of biologically vs clinically defined Alzheimer spectrum entities using the National Institute on Aging–Alzheimer's Association research framework. *JAMA neurology*, 76(10):1174–1183, 2019.

[87] Clifford R Jack Jr, David S Knopman, William J Jagust, Ronald C Petersen, Michael W Weiner, Paul S Aisen, Leslie M Shaw, Prashanthi Vemuri, Heather J Wiste, Stephen D Weigand, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology*, 12(2):207–216, 2013.

[88] Clifford R Jack Jr, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, and John Q Trojanowski. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, 9(1):119–128, 2010.

[89] Alastair Jamieson-Lane and Bernd Blasius. Calculation of epidemic arrival time distributions using branching processes. *Physical Review E*, 102(4):042301, 2020.

[90] Joseph T Jarrett and Peter T Lansbury Jr. Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in alzheimer's disease and scrapie? *Cell*, 73(6):1055–1058, 1993.

[91] M. Jenkinson, C. Beckmann, T. Behrens, M. Woolrich, and S. Smith. FSL. *NeuroImage*, 62:782–790, 2012.

[92] Keith A Johnson, Nick C Fox, Reisa A Sperling, and William E Klunk. Brain imaging in Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, 2(4):a006213, 2012.

[93] Keith A Johnson, Aaron Schultz, Rebecca A Betensky, J Alex Becker, Jorge Sepulcre, Dorene Rentz, Elizabeth Mormino, Jasmeer Chhatwal, Rebecca Amariglio, Kate Papp, et al. Tau positron emission tomographic imaging in aging and early a lzheimer disease. *Annals of neurology*, 79(1):110–119, 2016.

[94] Mathias Jucker and Lary C Walker. Pathogenic protein seeding in Alzheimer disease and other neurodegenerative disorders. *Annals of Neurology*, 70(4):532–540, 2011.

[95] Mathias Jucker and Lary C Walker. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature*, 501(7465):45–51, 2013.

[96] Mathias Jucker and Lary C Walker. Propagation and spread of pathogenic protein assemblies in neurodegenerative diseases. *Nature Neuroscience*, 21(10):1341–1349, 2018.

[97] Fuyuki Kametani and Masato Hasegawa. Reconsideration of amyloid hypothesis and tau hypothesis in alzheimer's disease. *Frontiers in neuroscience*, page 25, 2018.

[98] Jie Kang, Hans-Georg Lemaire, Axel Unterbeck, J Michael Salbaum, Colin L Masters, Karl-Heinz Grzeschik, Gerd Multhaup, Konrad Beyreuther, and Benno Müller-Hill. The precursor of alzheimer's disease amyloid a4 protein resembles a cell-surface receptor. *Nature*, 325(6106):733–736, 1987.

[99] Jie Kang and Benno Müller-Hill. Differential splicing of alzheimer's disease amyloid a4 precursor rna in rat tissues: Prea4695 mrna is predominantly produced in rat and human brain. *Biochemical and biophysical research communications*, 166(3):1192–1200, 1990.

[100] M.G. Kendall. A New Measure of Rank Correlation. *Biometrika*, 30(1):81–93, 1938.

[101] Csaba Kerepesi, Balázs Szalkai, Bálint Varga, and Vince Grolmusz. How to direct the edges of the connectomes: Dynamics of the consensus connectomes and the development of the connections in the human brain. *Plos One*, 11(6):e0158680, 2016.

[102] Csaba Kerepesi, Balazs Szalkai, Balint Varga, and Vince Grolmusz. The braingraph.org database of high resolution structural connectomes and the brain graph tools. *Cognitive Neurodynamics*, 11(5):483–486, 2017.

[103] Csaba Kerepesi, Balázs Szalkai, Bálint Varga, and Vince Grolmusz. Comparative connectomics: Mapping the inter-individual variability of connections within the regions of the human brain. *Neuroscience Letters*, 662:17–21, 2018.

[104] PG Kevrekidis, Travis B Thompson, and Alain Goriely. Anisotropic diffusion and traveling waves of toxic proteins in neurodegenerative diseases. *Physics Letters A*, 384(36):126935, 2020.

[105] Michael Kidd. Paired helical filaments in electron microscopy of alzheimer's disease. *Nature*, 197(4863):192–193, 1963.

[106] Hyunjeong Kim, Eosu Kim, Minsun Park, Eun Lee, and Kee Namkoong. Organotypic hippocampal slice culture from the adult mouse brain: a versatile tool for translational neuropsychopharmacology. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 41:36–43, 2013.

[107] Jungsu Kim, Luisa Onstead, Suzanne Randle, Robert Price, Lisa Smithson, Craig Zwizinski, Dennis W Dickson, Todd Golde, and Eileen McGowan. Aβ40 inhibits amyloid deposition in vivo. *Journal of Neuroscience*, 27(3):627–633, 2007.

[108] Nobuya Kitaguchi, Yasuyuki Takahashi, Yasuo Tokushima, Satoshi Shiojiri, and Hirataka Ito. Novel precursor of alzheimer's disease amyloid protein shows protease inhibitory activity. *Nature*, 331(6156):530–532, 1988.

[109] Sven Konzack, Edda Thies, Alexander Marx, Eva-Maria Mandelkow, and Eckhard Mandelkow. Swimming against the tide: mobility of the microtubule-associated protein tau in neurons. *Journal of Neuroscience*, 27(37):9916–9927, 2007.

[110] Christian Ledig, Andreas Schuh, Ricardo Guerrero, Rolf A Heckemann, and Daniel Rueckert. Structural brain imaging in Alzheimer's disease and mild cognitive impairment: biomarker analysis and shared morphometry database. *Scientific Reports*, 8(1):1–16, 2018.

[111] Antoine Leuzy, Konstantinos Chiotis, Laetitia Lemoine, Per-Göran Gillberg, Ove Almkvist, Elena Rodriguez-Vieitez, and Agneta Nordberg. Tau pet imaging in neurodegenerative tauopathies—still a challenge. *Molecular psychiatry*, 24(8):1112–1134, 2019.

[112] Tommy L Lewis Jr, Julien Courchet, and Franck Polleux. Cell biology in neuroscience: Cellular and molecular mechanisms underlying axon formation, growth, and branching. *The Journal of Cell Biology*, 202(6):837, 2013.

[113] Yi Li, Juha O Rinne, Lisa Mosconi, Elizabeth Pirraglia, Henry Rusinek, Susan DeSanti, Nina Kemppainen, Kjell Någren, Byeong-Chae Kim, Wai Tsui, et al. Regional analysis of fdg and pib-pet images in normal aging, mild cognitive impairment, and alzheimer's disease. *European journal of nuclear medicine and molecular imaging*, 35(12):2169–2181, 2008.

[114] 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.

[115] DMA Mann, A Brown, D Prinja, CA Davies, M Landon, CL Masters, and K Beyreuther. An analysis of the morphology of senile plaques in down's syndrome patients of different ages using immunocytochemical and lectin histochemical techniques. *Neuropathology and applied neurobiology*, 15(4):317–329, 1989.

[116] Joanna Masel, Vincent AA Jansen, and Martin A Nowak. Quantifying the kinetic parameters of prion replication. *Biophysical chemistry*, 77(2-3):139–152, 1999.

[117] Niklas Mattsson, Michael Schöll, Olof Strandberg, Ruben Smith, Sebastian Palmqvist, Philip S Insel, Douglas Hägerström, Tomas Ohlsson, Henrik Zetterberg, Jonas Jögi, et al. 18f-av-1451 and csf t-tau and p-tau as biomarkers in alzheimer's disease. *EMBO molecular medicine*, 9(9):1212–1223, 2017.

[118] Jorge F Mejias and Xiao-Jing Wang. Mechanisms of distributed working memory in a large-scale network of macaque neocortex. *Elife*, 11:e72136, 2022.

[119] Russell Merris. Laplacian matrices of graphs: a survey. *Linear algebra and its applications*, 197:143–176, 1994.

[120] Chris Mezias, Eve LoCastro, Chuying Xia, and Ashish Raj. Connectivity, not region-intrinsic properties, predicts regional vulnerability to progressive tau pathology in mouse models of disease. *Acta Neuropathologica Communications*, 5(1):61, 2017.

[121] E Mohandas, V Rajmohan, and B Raghunath. Neurobiology of alzheimer's disease. *Indian journal of psychiatry*, 51(1):55–61, 2009.

[122] Dunja Mrdjen, Edward J Fox, Syed A Bukhari, Kathleen S Montine, Sean C Bendall, and Thomas J Montine. The basis of cellular and regional vulnerability in Alzheimer's disease. *Acta Neuropathologica*, pages 1–21, 2019.

[123] Amritpal Mudher and Simon Lovestone. Alzheimer's disease–do tauists and baptists finally shake hands? *Trends in neurosciences*, 25(1):22–26, 2002.

[124] Satoshi Muraoka, Annina M DeLeo, Manveen K Sethi, Kayo Yukawa-Takamatsu, Zijian Yang, Jina Ko, John D Hogan, Zhi Ruan, Yang You, Yuzhi Wang, et al. Proteomic and biological profiling of extracellular vesicles from alzheimer's disease human brain tissues. *Alzheimer's & Dementia*, 16(6):896–907, 2020.

[125] Sangeeta Nath, Lotta Agholme, Firoz Roshan Kurudenkandy, Björn Granseth, Jan Marcusson, and Martin Hallbeck. Spreading of neurodegenerative pathology via neuron-to-neuron transmission of β-amyloid. *Journal of Neuroscience*, 32(26):8767–8777, 2012.

[126] M. E. J. (Mark E. J.) Newman. *Networks an introduction*. Oxford University Press, Oxford, 2010.

[127] K Ogomori, T Kitamoto, J Tateishi, Y Sato, M Suetsugu, and M Abe. Beta-protein amyloid is widely distributed in the central nervous system of patients with alzheimer's disease. *The American journal of pathology*, 134(2):243, 1989.

[128] Nobuyuki Okamura and Kazuhiko Yanai. Applications of tau pet imaging. *Nature Reviews Neurology*, 13(4):197–198, 2017.

[129] M Öst, K Nylen, L Csajbok, A Olsson Öhrfelt, M Tullberg, C Wikkelsö, P Nellgård, L Rosengren, K Blennow, and B Nellgård. Initial csf total tau correlates with 1-year outcome in patients with traumatic brain injury. *Neurology*, 67(9):1600–1604, 2006.

[130] Sneha Pandya, Amy Kuceyeski, and Ashish Raj. The brain's structural connectome mediates the relationship between regional neuroimaging biomarkers in Alzheimer's disease. *Journal of Alzheimer's Disease*, 55(4):1639–1657, 2017.

[131] Sneha Pandya, Chris Mezias, and Ashish Raj. Predictive model of spread of progressive supranuclear palsy using directional network diffusion. *Frontiers in Neurology*, 8:692, 2017.

[132] Virenkumar A Pandya and Rickie Patani. Region-specific vulnerability in neurodegeneration: Lessons from normal ageing. *Ageing Research Reviews*, 67:101311, 2021.

[133] Hugh A Pearson and Chris Peers. Physiological roles for amyloid β peptides. *The Journal of physiology*, 575(1):5–10, 2006.

[134] R Ca Pearson, MM Esiri, RW Hiorns, GK Wilcock, and TP Powell. Anatomical correlates of the distribution of the pathological changes in the neocortex in alzheimer disease. *Proceedings of the National Academy of Sciences*, 82(13):4531–4534, 1985.

[135] Rocío Pérez-González, Yohan Kim, Chelsea Miller, Javier Pacheco-Quinto, Elizabeth A Eckman, and Efrat Levy. Extracellular vesicles: where the amyloid precursor protein carboxyl-terminal fragments accumulate and amyloid-β oligomerizes. *The FASEB Journal*, 34(9):12922–12931, 2020.

[136] Lawrence Perko. *Differential equations and dynamical systems*, volume 7. Springer Science & Business Media, 2013.

[137] Camille Pernègre, Antoine Duquette, and Nicole Leclerc. Tau secretion: good and bad for neurons. *Frontiers in Neuroscience*, 13:649, 2019.

[138] Amy M Pooler, Emma C Phillips, Dawn HW Lau, Wendy Noble, and Diane P Hanger. Physiological release of endogenous tau is stimulated by neuronal activity. *EMBO reports*, 14(4):389–394, 2013.

[139] Parvoneh Poorkaj, Thomas D Bird, Ellen Wijsman, Ellen Nemens, Ralph M Garruto, Leojean Anderson, Athena Andreadis, Wigbert C Wiederholt, Murray Raskind, and Gerard D Schellenberg. Tau is a candidate gene for chromosome 17 frontotemporal dementia. *Annals of neurology*, 43(6):815–825, 1998.

[140] Thorsten Pöschel, Nikolai V Brilliantov, and Cornelius Frömmel. Kinetics of prion growth. *Biophysical journal*, 85(6):3460–3474, 2003.

[141] Stanley B Prusiner. Molecular biology of prion diseases. *Science*, 252(5012):1515–1522, 1991.

[142] Stanley B Prusiner. Prions. *Proceedings of the National Academy of Sciences*, 95(23):13363–13383, 1998.

[143] Prama Putra, Hadrien Oliveri, Travis Thompson, and Alain Goriely. Front propagation and arrival times in networks with application to neurodegenerative diseases. *SIAM Journal on Applied Mathematics*, 83(1):194–224, 2023.

[144] 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):929–956, 2021.

[145] Ashish Raj, Amy Kuceyeski, and Michael Weiner. A network diffusion model of disease progression in dementia. *Neuron*, 73(6):1204–1215, 2012.

[146] David Riascos, Dianne De Leon, Alaina Baker-Nigh, Alexander Nicholas, Rustam Yukhananov, Jing Bu, Chuang-Kuo Wu, and Changiz Geula. Age-related loss of calcium buffering and selective neuronal vulnerability in alzheimer's disease. *Acta neuropathologica*, 122:565–576, 2011.

[147] Baden Rumble, Robert Retallack, Caroline Hilbich, Gail Simms, Gerd Multhaup, Ralph Martins, Athel Hockey, Philip Montgomery, Konrad Beyreuther, and Colin L Masters. Amyloid a4 protein and its precursor in down's syndrome and alzheimer's disease. *New England Journal of Medicine*, 320(22):1446–1452, 1989.

[148] Laure Saint-Aubert, Laetitia Lemoine, Konstantinos Chiotis, Antoine Leuzy, Elena Rodriguez-Vieitez, and Agneta Nordberg. Tau pet imaging: present and future directions. *Molecular neurodegeneration*, 12(1):1–21, 2017.

[149] Kajsa Sämgård, Henrik Zetterberg, Kaj Blennow, Oskar Hansson, Lennart Minthon, and Elisabet Londos. Cerebrospinal fluid total tau as a marker of alzheimer's disease intensity. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, 25(4):403–410, 2010.

[150] Amelie Schäfer, Elizabeth C Mormino, and Ellen Kuhl. Network diffusion modeling explains longitudinal tau pet data. *Frontiers in Neuroscience*, 14:566876, 2020.

[151] Michael Schöll, Samuel N Lockhart, Daniel R Schonhaut, James P O'Neil, Mustafa Janabi, Rik Ossenkoppele, Suzanne L Baker, Jacob W Vogel, Jamie Faria, Henry D Schwimmer, et al. Pet imaging of tau deposition in the aging human brain. *Neuron*, 89(5):971–982, 2016.

[152] 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.

[153] Adam J Schwarz, Peng Yu, Bradley B Miller, Sergey Shcherbinin, James Dickson, Michael Navitsky, Abhinay D Joshi, Michael D Devous Sr, and Mark S Mintun. Regional profiles of the candidate tau pet ligand 18 f-av-1451 recapitulate key features of braak histopathological stages. *Brain*, 139(5):1539–1550, 2016.

[154] William W Seeley, Richard K Crawford, Juan Zhou, Bruce L Miller, and Michael D Greicius. Neurodegenerative diseases target large-scale human brain networks. *Neuron*, 62(1):42–52, 2009.

[155] M. Serrano, M. Boguna, and A. Vespignani. Extracting the multiscale backbone of complex weighted networks. *PNAS*, 106(16):6483–6488, 2009.

[156] Vivek Singh, Howard Chertkow, Jason P Lerch, Alan C Evans, Adrienne E Dorr, and Noor Jehan Kabani. Spatial patterns of cortical thinning in mild cognitive impairment and alzheimer's disease. *Brain*, 129(11):2885–2893, 2006.

[157] P.B. Slater. A two-stage algorithm for extracting the multiscale backbone of complex weighted networks. *PNAS*, 106(26):E66, 2009.

[158] Ruben Smith, Michael Schöll, Antoine Leuzy, Jonas Jögi, Tomas Ohlsson, Olof Strandberg, and Oskar Hansson. Head-to-head comparison of tau positron emission tomography tracers [18 f] flortaucipir and [18 f] ro948. *European journal of nuclear medicine and molecular imaging*, 47:342–354, 2020.

[159] S. Smith, M. Jenkinson, M. Woolrich, P. Matthews, et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23(S1):208–219, 2004.

[160] James Sonne, Vamsi Reddy, and Morris R. Beato. *Neuroanatomy, Substantia Nigra*. StatPearls Publishing, Treasure Island (FL), 2022.

[161] Claudio Soto. In vivo spreading of tau pathology. *Neuron*, 73(4):621–623, 2012.

[162] Maria Grazia Spillantini, R Anthony Crowther, Ross Jakes, Masato Hasegawa, and Michel Goedert. α-synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with Lewy bodies. *Proceedings of the National Academy of Sciences*, 95(11):6469–6473, 1998.

[163] Maria Grazia Spillantini, Marie Luise Schmidt, Virginia M-Y Lee, John Q Trojanowski, Ross Jakes, and Michel Goedert. α-synuclein in lewy bodies. *Nature*, 388(6645):839–840, 1997.

[164] Dorothea Strozyk, K Blennow, LR White, and LJ Launer. Csf aβ 42 levels correlate with amyloid-neuropathology in a population-based autopsy study. *Neurology*, 60(4):652–656, 2003.

[165] Balázs Szalkai, Csaba Kerepesi, Bálint Varga, and Vince Grolmusz. High-resolution directed human connectomes and the consensus connectome dynamics. *PloS One*, 14(4):e0215473, 2019.

[166] Balázs Szalkai, Bálint Varga, and Vince Grolmusz. Graph theoretical analysis reveals: Women's brains are better connected than men's. *PLoS One*, 10(7):e0130045, 2015.

[167] Ruensern Tan, Aileen J Lam, Tracy Tan, Jisoo Han, Dan W Nowakowski, Michael Vershinin, Sergi Simó, Kassandra M Ori-McKenney, and Richard J McKenney. Microtubules gate tau condensation to spatially regulate microtubule functions. *Nature cell biology*, 21(9):1078–1085, 2019.

[168] Tero Tapiola, Irina Alafuzoff, Sanna-Kaisa Herukka, Laura Parkkinen, Päivi Hartikainen, Hilkka Soininen, and Tuula Pirttilä. Cerebrospinal fluid β-amyloid 42 and tau proteins as biomarkers of alzheimer-type pathologic changes in the brain. *Archives of neurology*, 66(3):382–389, 2009.

[169] Mara Ten Kate, Silvia Ingala, Adam J Schwarz, Nick C Fox, Gaël Chételat, Bart NM van Berckel, Michael Ewers, Christopher Foley, Juan Domingo Gispert, Derek Hill, et al. Secondary prevention of Alzheimer's dementia: neuroimaging contributions. *Alzheimer's Research & Therapy*, 10(1):112, 2018.

[170] Dietmar R Thal, Udo Rüb, Mario Orantes, and Heiko Braak. Phases of aβ-deposition in the human brain and its relevance for the development of ad. *Neurology*, 58(12):1791–1800, 2002.

[171] Dietmar Rudolf Thal, Johannes Attems, and Michael Ewers. Spreading of amyloid, tau, and microvascular pathology in Alzheimer's disease: findings from neuropathological and neuroimaging studies. *Journal of Alzheimer's Disease*, 42(s4):S421–S429, 2014.

[172] 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.

[173] Justin Torok, Chaitali Anand, Parul Verma, and Ashish Raj. Connectome-based biophysics models of alzheimer's disease diagnosis and prognosis. *Translational Research*, 254:13–23, 2023.

[174] Justin Torok, Pedro D Maia, Parul Verma, Christopher Mezias, and Ashish Raj. Emergence of directional bias in tau deposition from axonal transport dynamics. *PLoS computational biology*, 17(7):e1009258, 2021.

[175] J.-D. Tournier, F. Calamante, and A. Connelly. MRtrix: Diffusion tractography in crossing fiber regions. *Int. J. Imaging Syst. Technol.*, 22(1):53–66, 2012.

[176] Neurodegenerative diseases and dementia. `https://mrc.ukri.org/funding/science-areas/neurosciences-mental-health/our-science-and-contacts-nmhb/neurodegenerative-diseases-and-dementia/`. Accessed: 2020-09-16.

[177] Spotlight on: dementia. `https://mrc.ukri.org/research/spotlights/dementia/`. Accessed: 2020-09-16.

[178] Victor L Villemagne, Vincent Doré, Samantha C Burnham, Colin L Masters, and Christopher C Rowe. Imaging tau and amyloid-β proteinopathies in alzheimer disease and other conditions. *Nature Reviews Neurology*, 14(4):225–236, 2018.

[179] Victor L Villemagne, Catriona A McLean, Katrina Reardon, Alison Boyd, Victoria Lewis, Genevieve Klug, Gareth Jones, David Baxendale, Colin L Masters,

Christopher C Rowe, et al. 11c-pib pet studies in typical sporadic creutzfeldt–jakob disease. *Journal of Neurology, Neurosurgery & Psychiatry*, 80(9):998–1001, 2009.

[180] Jacob W Vogel, Yasser Iturria-Medina, Olof T Strandberg, Ruben Smith, Elizabeth Levitis, Alan C Evans, and Oskar Hansson. Spread of pathological tau proteins through communicating neurons in human Alzheimer's disease. *Nature Communications*, 11(1):1–15, 2020.

[181] 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.

[182] Lary C Walker, Marc I Diamond, Karen E Duff, and Bradley T Hyman. Mechanisms of protein seeding in neurodegenerative diseases. *JAMA neurology*, 70(3):304–310, 2013.

[183] Lary C Walker and Mathias Jucker. Amyloid by default. *Nature neuroscience*, 14(6):669–670, 2011.

[184] Lary C Walker and Mathias Jucker. Neurodegenerative diseases: expanding the prion concept. *Annual Review of Neuroscience*, 38:87–103, 2015.

[185] Maxwell B Wang, Julia P Owen, Pratik Mukherjee, and Ashish Raj. Brain network eigenmodes provide a robust and compact representation of the structural connectome in health and disease. *PLoS Computational Biology*, 13(6), 2017.

[186] Joel C Watts, Carlo Condello, Jan Stöhr, Abby Oehler, Joanne Lee, Stephen J DeArmond, Lars Lannfelt, Martin Ingelsson, Kurt Giles, and Stanley B Prusiner. Serial propagation of distinct strains of aβ prions from alzheimer's disease patients. *Proceedings of the National Academy of Sciences*, 111(28):10323–10328, 2014.

[187] Johannes Weickenmeier, Mathias Jucker, Alain Goriely, and Ellen Kuhl. A physics-based model explains the prion-like features of neurodegeneration in Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis. *Journal of the Mechanics and Physics of Solids*, 124:264–281, 2019.

[188] Johannes Weickenmeier, Ellen Kuhl, and Alain Goriely. Multiphysics of prion-like diseases: Progression and atrophy. *Physical Review Letters*, 121(15):158101, 2018.

[189] M. Woolrich, S. Jbabdi, B. Patenaude, S. Smith, et al. Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, 45:S173–86, 2009.

[190] Jessica W Wu, Mathieu Herman, Li Liu, Sabrina Simoes, Christopher M Acker, Helen Figueroa, Joshua I Steinberg, Martin Margittai, Rakez Kayed, Chiara Zurzolo, et al. Small misfolded tau species are internalized via bulk endocytosis and anterogradely and retrogradely transported in neurons. *Journal of Biological Chemistry*, 288(3):1856–1870, 2013.

[191] Kaoru Yamada. Extracellular tau and its potential role in the propagation of tau pathology. *Frontiers in neuroscience*, 11:667, 2017.

[192] Kaoru Yamada, Tirth K Patel, Katja Hochgräfe, Thomas E Mahan, Hong Jiang, Floy R Stewart, Eva-Maria Mandelkow, and David M Holtzman. Analysis of in vivo turnover of tau in a mouse model of tauopathy. *Molecular neurodegeneration*, 10(1):1–9, 2015.

[193] H Yamaguchi, Y Nakazato, M Shoji, K Okamoto, Y Ihara, M Morimatsu, and S Hirai. Secondary deposition of beta amyloid within extracellular neurofibrillary tangles in alzheimer-type dementia. *The American journal of pathology*, 138(3):699, 1991.

[194] Haruyasu Yamaguchi, Koji Ishiguro, Mikio Shoji, Tsuneo Yamazaki, Yoichi Nakazato, Yasuo Ihara, and Shunsaku Hirai. Amyloid β/a4 protein precursor is bound to neurofibrillary tangles in alzheimer-type dementia. *Brain research*, 537(1-2):318–322, 1990.

[195] Azar Zandifar, Vladimir S Fonov, Simon Ducharme, Sylvie Belleville, D Louis Collins, Alzheimer's Disease Neuroimaging Initiative, et al. Mri and cognitive scores complement each other to accurately predict alzheimer's dementia 2 to 7 years before clinical onset. *NeuroImage: Clinical*, 25:102121, 2020.

[196] Hans Zempel and Eckhard Mandelkow. Lost after translation: missorting of tau protein and consequences for alzheimer disease. *Trends in neurosciences*, 37(12):721–732, 2014.

[197] Xiaozhu Zhang, Dirk Witthaut, Marc Timme, et al. Topological determinants of perturbation spreading in networks. *Physical Review Letters*, 125(21):218301, 2020.

[198] Ying-Qiu Zheng, Yu Zhang, Yvonne Yau, Yashar Zeighami, Kevin Larcher, Bratislav Misic, and Alain Dagher. Local vulnerability and global connectivity jointly shape neurodegenerative disease propagation. *PLoS biology*, 17(11):e3000495, 2019.

[199] 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.