Recovering Disease Trajectory in Neurodegeneration: An Agent-Based Approach

Ying-Qiu Zheng

Integrated Program in Neuroscience,

McGill University, Montreal

April 2019

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science.

Contents

1 Abstract

It is becoming increasingly clear that brain network organization shapes the course and expression of neurodegenerative diseases. Parkinson's disease (PD) is marked by progressive spread of atrophy from the midbrain to subcortical structures and, eventually, to the cerebral cortex. Recent discoveries suggest that the neurodegenerative process involves the misfolding of endogenous proteins (α-synuclein) and prion-like spread of these pathogenic proteins via axonal projections. However, the mechanisms that translate local "synucleinopathy" to large-scale network dysfunction and atrophy remain unknown. Here we use an agent-based epidemic spreading model to integrate structural connectivity, functional connectivity and gene expression, and to predict sequential volume loss due to neurodegeneration. We demonstrate four key findings. First, the dynamic model replicates the spatial distribution of empirical atrophy identified in an independent dataset of PD patients. Second, the model implicates the substantia nigra as the disease epicenter, consistent with previous literature. Third, we reveal a significant role for both connectome topology and spatial embedding (geometry) in shaping the distribution of atrophy. Finally, regional gene expression and functional co-activation further amplify the course set by connectome architecture. Altogether, these results support the notion that the progression of neurodegenerative disease is a multifactorial process that depends on both cell-to-cell spreading of misfolded proteins and regional vulnerability. The model proves powerful in modelling neurodegeneration and provides insights into developing and testing preventative procedures.

2 Resume

Il devient de plus en plus évident que l'organisation des réseaux cérébraux détermine l'évolution et l'expression des maladies neurodégénératives. La maladie de Parkinson (MP) se caractérise par une augmentation progressive de l'atrophie du mésencéphale aux structures sous-corticales et, éventuellement, au cortex cérébral. De récentes découvertes suggèrent que le processus neurodégénératif implique un repliement anormal des protéines endogènes (alpha-synucléine) et que la propagation (analogue au prion) de ces protéines pathogènes se ferait via les projections axonales. Cependant, les mécanismes impliqués dans l'évolution de la synucléinopathie locale en dysfonctionnement et atrophie du réseau à grande échelle restent inconnus. Dans cette étude, nous avons utilisé un modèle de propagation épidémique basé sur les agents pour intégrer la connectivité structurelle, la connectivité fonctionnelle et l'expression génétique ainsi que pour prédire la perte de volume séquentielle causée par la neurodégénérescence. Nous avons démontré quatre principaux résultats. Premièrement, le modèle dynamique reproduit la distribution spatiale de l'atrophie empirique identifiée dans une base de données indépendante de patients atteints de la MP. Deuxièmement, conformément à la littérature antérieure, le modèle implique la substance noire comme épicentre de la maladie. Troisièmement, nous avons révélé que la topologie du connectome et que l'intégration spatiale (la géométrie) joue un rôle significatif dans la distribution de l'atrophie. Finalement, l'expression régionale génétique et la co-activation fonctionnelle amplifient davantage le cours défini par l'architecture du connectome. Globalement, ces résultats appuient la notion selon laquelle la progression des maladies neurodégénératives est un processus multifactoriel qui dépend à la fois de la propagation des protéines avec un repliement anormal et de la vulnérabilité régionale. Le modèle s'avère efficace pour la modélisation de la neurodégénérescence et il fournit des informations importantes pour développer et tester des procédures préventives.

3 Acknowledgement

First of all I would like to express my sincere thanks to my supervisors Dr. Alain Dagher and Dr. Bratislav Mišić who have supervised the work and provided immense help with study design, methodological development and manuscript construction, and by no means the least their kind emotional support on my scientific career. My thesis work would not have been possible without their patient guidance and invaluable assistance. I also wish to thank my committee member Dr Ziv Or-Gan for providing many extremely useful suggestions.

I would like to thank my friends and colleagues at the Dagher lab for creating a homely and stimulating environment. My co-authors, Yu Zhang, Yvonne Yau, Yashar Zeighami and Kevin Larcher, have provided huge help in data processing and manuscript preparation. I am very grateful to Dr. Christina Tremblay for providing a nice French translation of the abstract.

I would like to thank the multiple funding agencies for providing the financial support, including the Canada First Research Excellence Fund, the Canadian Institutes for Health Research, Natural Sciences and Engineering Research Council of Canada, Michael J Fox Foundation, Weston Brain Institute, and the Alzheimer Association.

4 Author Contributions

I have performed the major methodological development and model implementations, data processing and results analysis. Dr. Alain Dagher and Dr. Bratislav Mišić have conceived, supervised and help constructed the work. The contributions of co-authors include data processing, manuscript revision and technical discussions.

5 Introduction

The main purpose of this thesis is to (1) propose a new approach to model the propagation of proteinopathy in neurodegeneration and (2) understand how ‘pathogenic spread’ and ‘selective neuronal vulnerability’, the two principal mechanisms of neurodegeneration, collectively shape the pathology of neurodegeneration under the modulation of brain network and gene epressions. The motivation, review of previous literature, scientific contributions and methodology are presented in the following sections.

5.1 Motivation

Neurodegenerative diseases such as Alzheimer's Disease (AD), Parkinson's Disease (PD), and Amyotrophic Lateral Sclerosis, are a major cause of psychosocial burden and mortality, but lack specific therapy. Until recently, the mechanism of progressive neuronal death in these conditions was unknown. However, converging lines of evidence from molecular, animal and human postmortem studies point to misfolded neurotoxic proteins that propagate through the central nervous system via neuronal connections (Brundin and Melki, 2017; Guo and Lee, 2014; Polymenidou and Cleveland, 2012; Jucker and Walker, 2013; Brettschneider et al., 2015; Walsh and Selkoe, 2016). These pathogenic misfolded disease-specific proteins function as corruptive templates that induce their normal protein counterparts to adopt a similar conformational alteration, analogous to the self-replication process in prion diseases. Examples include amyloid β-protein (Aβ) and *tau* in AD and α-synuclein in PD. The misfolded proteins can deposit into insoluble aggregates and progressively spread to interconnected neuronal populations through synaptic connections. The model of a propagating proteinopathy remains controversial however (Surmeier et al., 2017), and direct evidence in humans remains mostly circumstantial (Kordower et al., 2008). Moreover, other mechanisms may also drive PD pathology, including cell-autonomous factors, dependent on gene expression that modulate regional neuronal vulnerability (Surmeier et al., 2017). It would be timely and important to reassess the proteinopathy propagation hypothesis and investigate whether selective vulnerability may also modulates the pathology, and if yes, how it may collectively shape disease spread with the proteinopathy propagation mechanism. We propose to use computational models of neurodegeneration to investigate these questions, which may provide new insights into the pathological processes underlying disease progression, independent of evidence from animal studies. For example, if the pathology of neurodegenerative diseases is indeed driven by progressive accumulation and propagation of disease-related proteins, such a model should recapitulate the spatial pattern of regional neurodegeneration in patients thereby providing converging and independent evidence for the pathogenic spread hypothesis.

The developments of magnetic resonance imaging (MRI) and "big data" neuroimaging consortiums in recent decades provide an unprecedented opportunity to probe structural and functional abnormality of patients with neurodegeneration. For example, structural MRI provides measurement of local volume changes with respect to normal aging, and diffusion MRI makes it possible to infer the fiber density between brain regions. These techniques allows validating macroscopic consequences of disease spreading that involves microscopic physiology. Taking PD as an example, here we propose a model that integrates multimodal neuroimaging and genetic data to simulate the microscopic protein interaction and propagation of neurodegeneration and ask if the macroscopic consequence of proteinopathy propagation can recapitulate the spatial distribution of neuronal loss in patients. We will also incorporate local differences in gene expressions to test the influence of selective neuronal vulnerability.

5.2 Scientific contributions to original knowledge

The scientific contributions to original knowledge include:

- **We proposed a novel Susceptible-Infectious-Removed (S-I-R) agent-based model on a brain network to model the spreading of pathological proteins in neurodegenerative diseases (FIG. 1).** The agents are individual proteins. The population is split into S, the portion yet to be infected (normal proteins); I, the portion capable of transmitting the infection (misfolded proteins); and R, the portion no longer active in the spreading (metabolized and cleared proteins). We took PD as an example to show how an S-I-R agent-based model can track the spreading of misfolded α-synuclein, the pathological fibrillar species of endogenous α-synuclein suggested to be responsible for PD pathology.

- **We thoroughly investigated the two competing mechanisms, "pathogenic**

spread" and "selective neuronal vulnerability", in driving the progression of neurodegeneration. Brain network architecture (measured by fiber tracts density) governs the spreading of protein agents, providing a description of the non-cell-autonomous proteinopathy propagation, while gene expressions (here we focus on two genes, *SNCA* and *GBA*) encode protein synthesis and metabolism thereby modulating local variations in vulnerability. We showed that the spatial distribution of neuronal loss in PD results from the collective interplay between the two mechanisms.

- **We revealed how connectome architecture, gene expressions and neural activity shape disease spread.** After showing that the S-I-R agent-based model can recapitulate the atrophy pattern in PD patients, we destroyed the original information of its constituent factors, including disrupting the architecture of the connectome (physical routes of spreading), shuffling gene expressions (local variations in synuclein synthesis and clearance) and randomizing neural activity (bias the spreading process on top of the actual routes). We found all the procedures remarkably degraded model performance, suggesting their significant roles in driving disease progression.

5.3 Review of relevant literature

5.3.1 Modelling of neurodegeneration and epidemic models

Previous efforts to model neurodegeneration have primarily focused on diffusion models (Raj et al., 2012). In simple diffusion models, disease in any region is modeled as a concentration (e.g. of misfolded protein) and propagation obeys the law of mass effect with first order kinetics. The epidemic is deterministic, defined by a series of differential equations. They can be installed either on a network and solved analytically via the graph Laplacian (Raj et al., 2012; Pandya et al., 2019), or on the whole brain and solved using finite element analysis (Weickenmeier et al., 2018a). Such models are easily solved mathematically but have limited explanatory power. Moreover, diffusion models are based on the assumption that disease spread is primarily driven by concentration gradients (i.e., high concentration of the protein slows down disease spread), which is not biologically realistic for prion-like process, in which high concentration of infectious agents may speed up spreading.

The chain-reaction-like prion hypothesis suggests that methods derived from infectious disease epidemiology serve as a good candidate for modelling propagation dynamics in neurodegenerative diseases. Just as infectious diseases spread via social networks, misfolded proteins propagate via the brain's connectome. Various epidemic models have been proposed over decades, including S-I, S-I-R, S-I-S, S-E-I-R and their numerous variations (Newman, 2010), and epidemic processes on networks in which vertices (or nodes) of the graph are hosts that adopt different states of a disease and infection initiated at seeds spreads on a graph-like structure, such as cascade models (Watts, 2002) and percolation on networks (Moore and Newman, 2000). Epidemic models can take an agent-based setting, in which the infectious state of each individual agent and its motility are simulated, and where simple local interactions can translate into complex global behavior (Frias-Martinez et al., 2011). Agent-based models have the advantage of easily incorporating additional emergent properties of a system as the epidemic spreads – for example a brain region may lose its ability to propagate the disease once it is severely affected. They also easily incorporate differences among agents (e.g. in susceptibility to infection or mobility), and are useful for testing interventions (e.g. vaccination).

5.3.2 Hypotheses of PD progression

Emerging evidence from post-mortem(Braak et al., 2004; Goedert et al., 2013) and neuroimaging studies(Zeighami et al., 2015) in recent decades have indicated that the pathology in PD is not restricted to the SN and encompasses a multitude of brain regions. Postmorten analysis of patients at different stages of PD suggest a temporal sequence of regional Lewy inclusions (Braak et al., 2003, 2004). Therefore, PD has been widely suggested as a "spreading" disease, which is often associated with α-synuclein, an endogenous protein found primarily in terminals and expressed throughout the neocortex and various subcortical structures(Jucker and Walker, 2013). Growing studies at the molecular level have consistently indicated that α-synuclein can adopt a maladaptive conformation and spread to interconnected neurons, functioning as seeds or templates to induce further abnormal folding of non-pathological protein counterparts. Misfolded α-synuclein can assemble into toxic insoluble fibrillar inclusions and accumulate progressively across neuronal populations and brain regions, causing widespread neuronal dysfunction and eventually cell death (Mougenot et al., 2012; Brettschneider et al., 2015). This pathogenic spread hypothesis of misfolded α-synuclein has been put on firmer ground by a series of animal studies(Luk et al., 2012; Mougenot et al., 2012; Goedert et al., 2013; Masuda-Suzukake et al., 2013; Peelaerts et al., 2015; Rey et al., 2016; Surmeier et al., 2017) revealing that α-synuclein fibrils, when injected directly into the brain, propagate to physically connected, neighbouring structures, bringing about Lewy-like pathology. Direct evidence for the actual neuron-to-neuron spread and aggregation of abnormally folded α-synuclein occurs in human PD patients includes the identification of Lewy pathology in fetal tissue grafts over decades after transplatation (Kordower et al., 2008; Li et al., 2008; Brundin and Kordower, 2012). However, these observations do not exclude other explanations, for example, that the surrounding PD pathogenic processes facilitate the misfolding of the endogenous α-synuclein in the grafted cells rather than directly propagate misfolded α-synuclein into them (Walsh and Selkoe, 2016). Moreover, it remains unclear concerning whether the pathogenic spread of misfolded α-synuclein aggregates is required for neuronal dysfunction and cell death, as well as the temporal sequence of regional volume loss over decades. Therefore, focusing on the α-synuclein propagation hypothesis as the core mechanism underlying PD seems premature.

The main alternative explanation of the symptom progression in PD is selective neuronal vulnerability. Regional variation in vulnerability to the underlying pathogenic processes of α-synuclein deposits (Mattson and Magnus, 2006; Saxena and Caroni, 2011; Jackson, 2014; Walsh and Selkoe, 2016), probably driven by gene expression profiles, shapes the spatiotemporally organized pattern of regional lesion accrual over decades. However, neither the spread model (i.e., non-cell-autonomous) nor selective vulnerability (i.e., cell-autonomous) fit perfectly with the symptom progression in PD. It has been proposed that the non-cell-autonomous and cell-autonomous factors are not mutually exclusive, and rather, interactively determine PD progression (Walsh and Selkoe, 2016).

5.4 Overview of thesis

Using brain imaging data from Parkinson's Progression Markers Initiative (PPMI) (Marek et al., 2011) and Human Connectome Project (HCP) (Van Essen et al., 2013), and gene expression profiles from the Allen Human Brain Atlas (AHBA) (Hawrylycz et al., 2012), we constructed a S-I-R agent-based model to simulate the spreading of misfolded α-synuclein and regional accrual of neuronal loss (atrophy). Measurements of empirical atrophy in PD patients were derived from structural magnetic resonance imaging (MRI) scans in PPMI and healthy connectomes that characterize the mobility pattern of the agents (α-synuclein) were constructed from HCP, while regional gene expression of *SNCA* and *GBA*, which modulate synuclein formation and degradation, were derived from the AHBA. Simulated regional atrophy derives from (1) the accumulation of misfolded α-synuclein aggregates and (2) deafferentation due to neurodegeneration in interconnected neuronal populations.

The study is organized as follows: first, we determine whether the model replicates empirically measured brain atrophy in PD; second, we identify the likely epicenter for disease initiation; third, we assess whether the disease spread is driven by non-cell-autonomous factors (structural connectivity) and cell-autonomous vulnerability (regional gene expression data); finally, we assess the role of activity-dependent spread of α-synuclein by incorporating resting state functional connectivity information into the model.

6 Results

6.1 Model contruction

- *Structural connectivity.* Diffusion-weighted MRI data from N=1027 healthy participants was used to construct the anatomical network for α-synuclein propagation (source: HCP, 2017 S1200 release; (Van Essen et al., 2013)). Adjacency matrices were reconstructed using deterministic streamline tractography (Yeh et al., 2013). A group consensus structural connectivity matrix was constructed by selecting the most commonly occurring edges averaged across all subjects, resulting in a binary density of 35% (Mišić et al., 2015, 2018).

- *Functional connectivity.* Resting-state functional MRI (fMRI) data from N=496 healthy participants (source: HCP, 2015 S500 release; (Van Essen et al., 2013)) was used to construct the functional connectome. Individual functional connectivity matrices were calculated using Pearson's correlation coefficient and then normalized using Fisher's z transform. A group correlation matrix was then constructed by first averaging the z-score connectivity matrices across subjects, and then converted back to correlation values using the inverse transformation. Negative correlation values in the resultant group connectivity matrix were set to zero.

- *Gene expression.* mRNA transcription (measured using in-situ hybridization) profiles of *SNCA* and *GBA* were averaged across samples in the same brain parcel and across the six subjects in the AHBA dataset. These gene expression profiles determine the local synthesis and degradation of α-synuclein.

- *Atrophy.* An atrophy map was derived from T1-weighted MRI scans of 237 PD patients and 118 age-matched healthy controls (source: PPMI; (Marek et al., 2011). For each participant (patient or healthy control), the Deformation-based Morphometry (DBM) value in each parcel was estimated to quantify the local volume change, on which an un-paired t-test was conducted between the patients and healthy controls. The resulting t-statistics were taken as the measure of regional atrophy (Zeighami et al., 2015).

The brain MRI template was parcellated according to an anatomical segmentation-based atlas, featuring 68 bilateral cortical and 15 subcortical regions (Desikan et al., 2006; Cammoun et al., 2012; Keuken et al., 2014). As only two of the six post-mortem AHBA brains have right hemispheric data available, and diffusion tractography is prone to errors in detecting interhemispheric connections, we simulated propagation using only the left hemisphere from the model, yielding 42 regions in total.

Synuclein propagation. We posited that regional expression level of endogenous α-synuclein already existing in the brain before disease onset may bias the trajectory of misfolded α-synuclein propagation. Therefore, to estimate regional density of endogenous α-synuclein in healthy brain, we set up a process that used generic information only to simulate the population growth of normal α-synuclein agents. Normal agents in region i are synthesized in each unit area (1mm^3 voxel) per unit time with probability α_i (the synthesis rate in region i). Meanwhile, any agent already existing in region i can: (a) exit region i and move into the edges it connects to with probabilities proportional to the corresponding connection strengths (densities of the fiber tracts); (b) remain in region i where it may be metabolized with probability β_i (the clearance rate in region i). Likewise, the agents in edge (i, j) can (a) exit edge (i, j) to enter region j with probability $1/l_{ij}$ where l_{ij} is the mean length of the fiber tracts between region i and j, reflecting our intuition that agents in longer edges have lower probability of exiting the edge; (b) remain in edge (i, j) with probability $1 - 1/l_{ij}$. The synthesis rate α_i and clearance rate β_i in region i are the *SNCA* and *GBA* expression z-scores respectively in region i converted to $(0, 1)$ using the standard normal cumulative distribution function. The system has only one stable point which can be found numerically (see *Analysis of the Fixed Points* in *Appendices*), suggesting that the growth of α-synuclein will deterministically converge to an equilibrium state set by the connectome and the gene expression profiles. The regional density of normal agents (number of agents per voxel) solved at the stable point was taken as the initial state of the system on which to simulate the misfolded α-synuclein spreading process.

Synuclein misfolding. We next initiated the pathogenic spread by injecting misfolded α-synuclein agents into the seed region, here chosen as the substantia nigra. The updating rules of normal agents (above) were adapted to account for their susceptibility to infection from contact with misfolded agents. Apart from the rules defined in the aforementioned growth process, normal (susceptible) agents in region i that survive degradation can be infected with probability γ_i thereby becoming misfolded (infected) agents (*Appendices*). In the absence of any molecular evidence to the contrary, misfolded agents are updated with the same mobility (exiting/remaining in regions/edges) and degradation (clearance rate) as normal agents. The new system seeded with misfolded α-synuclein has two fixed points: (1) one represents the scenario in which misfolded α-synuclein dies out, cleared by metabolic mechanisms before being able to transmit the infection to the entire population; (2) the other represents a major outbreak of misfolded α-synuclein, spreading to other regions via physical connections, causing further misfolding of endogenous α-synuclein and widespread propagation (Fig. S1). In this model, neither the injection number of misfolded α-synuclein agents nor the choice of seed region will affect the magnitude of misfolded α-synuclein accumulation at the fixed point; rather, they determine whether the spreading process converges to the epidemic scenario or dies out quickly. See Table S1 for the full list of parameters and their explanations.

6.2 Simulated neuronal loss replicates the spatial pattern of atrophy

We first investigated whether misfolded α-synuclein spreading on the healthy connectome can replicate the spatial patterning of atrophy observed in PD patients. We simulated the propagation of misfolded agents and the accrual of atrophy due to the toxic accumulation of the aggregates. Two factors that may induce neuronal loss were accounted for: (1) the accumulation of misfolded α-synuclein, which will cause region-specific cell death directly;

(2) atrophy due to deafferentation secondary to cell death in connected regions. At each time point, we compared the relative magnitude of simulated atrophy with the spatial pattern of empirical atrophy using Spearman's rank correlation coefficient, yielding the model fit as a function of time t.

As the misfolded agents propagate and accumulate in the system, the model fit increases up to a maximum value (r=0.63, p=1.71×10^{-5}, Fig. 2(a)) after which it drops slightly and stabilizes. It is possible that the slight decrease following the peak occurs because simulated atrophy becomes increasingly widespread as the propagation of misfolded agents progresses, while the empirical atrophy was derived from *de novo* PD patients at their first-visit in PPMI. Fig. 2(b) shows the linear relationship between simulated and empirical atrophy across all nodes at peak fit, while Fig. 2(c) shows the spatial correspondence between simulated and empirical atrophy. Interestingly, the model fit finally stabilizes with increasing t as the regional accumulation of misfolded α-synuclein approximates the stable point (see Fig. S4 for model fit up to 10^5 time steps), a finding that mirrors recent discoveries in animal models that α-synuclein eventually ceases to propagate in later stages (Rey et al., 2018). We also note that misfolded α-synuclein arrival time at each brain region resembles the well-established Braak stages of PD (Braak et al., 2003, 2004) (Fig. S5).

We next investigated if the model fit was consistent across variations in structural network connection densities. We selected varying subsets of the most commonly occurring edges in the individual structural connectivity matrices, varying the binary density of the group structural network matrix from 25% to 45% (of all possible edges). We then simulated the spreading processes on each network, derived the neuronal loss at each region and compared it with the empirical atrophy pattern using Spearman's rank correlation coefficient. All the simulations yielded comparable model fits with the peak correlation values consistently around 0.6 (Fig. 3, blue curve), suggesting that the S-I-R agent-based model is robust to variations in network density. Notably, we also assessed the Spearman's correlation between the regional density of misfolded α-synuclein and the empirical atrophy pattern. Across the same set of networks, simulated atrophy consistently provides better fits with the empirical atrophy than the regional density of misfolded α-synuclein (Fig. 3, red curve), indicating that accounting for cell death due to both α-synuclein and deafferentation yields a better model of regional atrophy accrual than the mere accumulation of misfolded α-synuclein.

Finally, we investigated whether the observed atrophy patterns could be directly reproduced from simpler topological measures, without invoking agent-based dynamics. We first tested whether simple regional variation in *GBA* or *SNCA* expression is associated with regional variation in atrophy. Neither *GBA* nor *SNCA* expression profiles bear a strong association with the spatial map of empirical atrophy (*GBA*: Spearman's r=-0.2402, p=0.1301; Pearson's r=-0.3109, p=0.0478; *SNCA*: Spearman's r=-0.2385, p=0.1330; Pearson's r=-0.2824, p=0.0736). Next, we tested whether simple network metrics provide a comparable fit to the observed atrophy values. We correlated the atrophy map with node-level network metrics including node degree, node strength, and eigenvector centrality at each network density ranging from 25% to 45%. Hubs, or nodes with greater degree connectivity or centrality, tend to be more atrophied (Fig. 3, green, purple and yellow curves), echoing the findings that hubs are often implicated in a host of brain disorders (Crossley et al., 2014). However, none of the metrics performed as well as the full agent-based model in matching the spatial pattern of empirical atrophy. Altogether, these results suggest that the protein dynamics embodied by the S-I-R agent-based model provide predictive power above and beyond network topology and gene expression.

6.3 Identifying the disease epicenter

We next investigated whether the model yields a disease epicenter consistent with previous literature. In the aforementioned process of normal α-synuclein growth, we solved the regional density of normal agents at the stable point as a baseline estimation of endogenous α-synuclein level in healthy brains. Recent findings from animal studies have suggested that α-synuclein expression level correlates with neuronal vulnerability in PD (Rey et al., 2018;

Luna et al., 2018); likewise, in our model, higher regional abundance of normal α-synuclein agents indicates greater likelihood of exposure to and growth of infectious agents, higher chance of disease transmission, and consequently, greater vulnerability to the accumulation of misfolded α-synuclein.

We compared the regional density of normal α-synuclein agents with the empirical selective vulnerability in patients to identify if highly vulnerable regions, such as substantia nigra, also manifest abundance of α-synuclein. We find that, of the 42 left hemisphere regions, substantia nigra has the highest normal α-synuclein level (Fig. 4, blue line). The elevated density of endogenous α-synuclein renders substantia nigra susceptible to the encroaching of infectious misfolded α-synuclein in the model, increasing both its vulnerability to misfolded protein and its chance of acting as a disease epicenter to further the propagation of the epidemic. This corresponds with the clinical observations of Lewy body inclusions and dopaminergic neuron loss identified in substantia nigra of PD patients as well as its role in most of the presenting symptoms of the disease (Spillantini et al., 1998; Damier et al., 1999; Braak et al., 2004). Moreover, other basal ganglia regions have relatively high levels of normal α-synuclein at the equilibrium compared to cortical regions (caudate ranks among the highest 42.9% of all the regions; putamen, 31.0%; pallidum, 28.6%), consistent with their role in propagating the disease process to the cerebral cortex (Yau et al., 2018). These findings suggest that our model can indeed represent regional variations in selective vulnerability to the pathogenic attacks underlying PD progression by combining information from the healthy connectome and *SNCA* and *GBA* expressions.

An alternative definition of disease epicenter is the seed node most likely to initiate a disease outbreak. As explained in the previous section, the agent-based model has two fixed points representing disease extinction or major outbreak. Although the choice of seed region and injection number of misfolded α-synuclein agents does not affect the magnitude of misfolded α-synuclein accumulation, it can initially shift the properties of the two fixed points, determining which one the system will converge to. We posited that the probability of triggering an outbreak indicates the plausibility of acting as an epicenter. Therefore, we quantified the spread threshold for each seed region, i.e., the minimally-required injection amount of misfolded α-synuclein to initiate an outbreak. In traditional epidemic disease models that do not consider spatial structure or synthesis of new susceptible hosts, *basic reproduction number* R_0 (the average number of susceptible agents that could be affected by an infectious agent before it has been removed) marks the transition between the regimes in which disease spreads or extinguishes (Newman, 2010). However, in our agent-based higher-order system in which new agents are constantly synthesized and move across regions, the transition threshold can only be determined numerically.

Substantia nigra has the lowest spread threshold (Fig. 4, red line), suggesting that it is also the most plausible seed region to initiate an epidemic spread. This is consistent with the notion that substantia nigra acts as the epicenter for propagation to the supratentorial central nervous system (Zeighami et al., 2015), and is generally one of the earliest regions to display neuronal loss in PD. Interestingly, other basal ganglia regions also exhibited relatively low spread thresholds (caudate ranks among the lowest 35.7% of all the regions; putamen, 38.1%; pallidum, 16.7%). Note however that our model does not include regions caudal to the midbrain (see *Discussion*).

6.4 Connectome architecture shapes disease spread

We next asked whether model fit would be facilitated or degraded by disrupting the connectome's topology or spatial embedding (geometry). To address this question, we implemented two types of null models, in which (a) the topology of the connectome was randomized (rewired null); or (b) the spatial positions of the regions were shuffled (spatial null).

Rewired null networks were generated by swapping pairs of edges while preserving the original degree sequence and density using the Maslov-Sneppen algorithm (Maslov and Sneppen, 2002) implemented in the Brain Connectivity Toolbox (https://sites.google.com/site/bctnet/) (Rubinov and Sporns, 2010). Note that it is possible that the edges after swapping are not defined in the original connectivity matrices (because no actual fiber tracts exist between the two regions). To interpolate fiber length in the rewired null network, for each region pair (i, j), we calculated the Euclidean distances between every possible pair of voxels respectively belonging to region i and j, and took the median as the distance between region i and j. Next, we fitted a simple linear regression model on the originally existing edges (i.e., $y = w_0 + w_1 x + \epsilon$, where y is the fiber length and x is the distance defined as above) and assigned the predicted fiber lengths to the new connections created during the rewiring process. Spatial null networks were generated by swapping the physical positions of the nodes while keeping their original connection profiles (Roberts et al., 2016; Seguin et al., 2018). This null model retains the degree sequence and connection profiles of every region, but randomizes spatial proximity. Networks at binary density 25%, 30%, 35% and 40% were selected as representatives to construct the two types of null networks, with 10,000 realizations each. We then implemented the dynamic model on each network and compared model fits for the empirical and null networks.

The agent-based model simulated on top of the empirical structural network yielded significantly greater fit to empirical atrophy than models simulated on either type of null network. This result was consistent across network densities (rewired null, Fig. 5(a): $p_{25\%} < 0.001$, $p_{30\%} = 0.0035$, $p_{35\%} = 0.0013$, $p_{40\%} = 0.0035$; spatial null, FIG. 5(b): $p_{25\%} < 0.001$, $p_{30\%} < 0.001$, $p_{35\%} < 0.001$, $p_{40\%} < 0.001$) and suggests that the high correspondence between simulated and empirical atrophy in PD is jointly driven by connectome topology and geometry.

6.5 Gene expression shapes disease spread

We next sought to directly assess the influence of local gene expression on spreading patterns of neurodegeneration. Based on molecular evidence, the model uses regional expression of *GBA* and *SNCA* as determinants of α-synuclein clearance and synthesis rate (Fig. S6). (Note however that any other gene known to influence α-synuclein synthesis or dynamics could also be included in the model.) Regional *GBA* and *SNCA* expressions were shuffled 10,000 times respectively by re-assigning the expression scores in each parcel (Fig. 6(a),(b) respectively). We then implemented the dynamic models with randomized expression levels and compared differences in model fit when using the empirical gene expression levels (Fig. 6, red) and permuted gene expression levels (Fig. 6, grey). Note that although *GBA* is an important genetic risk factor of Parkinson's disease, the removal of α-synuclein is a complex process and it is still not clear what role *GBA* does play in α-synuclein physiology. Therefore, we also adopted *TMEM175* as a substitute for *GBA* to modulate the breakdown of α-synuclein and then tested if the shuffled *TMEM175* expressions could still yield the same model fit. *TMEM175* deficiency has been reported to impair lysosomal and mitochondrial function and increases α-synuclein aggregation (Jinn et al., 2017). We argue that if lysosomal function does play a major part in α-synuclein clearance, the model using *TMEM175* to modulate clearance rate will output the same level of model fit as in the *GBA* model, which will not be observed when randomized *TMEM175* expressions are used.

Shuffling the transcription profile of either gene significantly degraded model fit (Fig. 6(a), *GBA*: $p_{25\%} = 0.0031$, $p_{30\%} < 0.001$, $p_{35\%} < 0.001$, $p_{40\%} = 0.0024$; FIG. 6(b), *SNCA*: $p_{25\%} = 0.0102$, $p_{30\%} = 0.0201$, $p_{35\%} = 0.0084$, $p_{40\%} = 0.0334$) suggesting a significant role of *GBA* and *SNCA* expression in driving the spatial patterning of atrophy. In other words, the regional expression of the genes, as implemented in the dynamic model, serves to modulate the vulnerability of individual nodes above and beyond their topological attributes by influencing α-synuclein synthesis, seeding and clearance. Comparable results were found in the models using *TMEM175* as a substitute for lysosomal or *GBA* ($p_{25\%} = 0.0232$, $p_{30\%} = 0.0071$, $p_{35\%} = 0.0210$, $p_{40\%} = 0.0215$).

An alternative explanation for these results is that simply introducing regional heterogeneity in gene expression levels improves model fit, for example because of differences in general transcription levels between cortex and subcortex. To address this possibility, we further assessed model fit in the cases where *GBA* and *SNCA* expression is made uniform across all brain regions. Instead of using empirical gene expression, we set uniform synthesis/clearance rates across all regions using the mean expression score, converted to a scalar value between $[0,1]$ using the standard normal cumulative distribution function. We then computed the model fit (peak Spearman's correlation value) for this "uniform" model. The models using uniform transcription profiles underperformed compared to those using empirical transcription profiles (Fig. 6, red = empirical, blue = uniform); in other words, the incorporation of true local differences in gene expression improves model fit, suggesting that the atrophy pattern in PD is not solely explained by pathogenic spreading *per se* but also depends on local vulnerability. Models implemented using uniform transcription profiles of either gene exhibited above-chance model fit compared to shuffled transcription profiles (*GBA* uniform correlations: $r_{25\%} = 0.4479$, $r_{30\%} = 0.3869$, $r_{35\%} = 0.3672$, $r_{40\%} = 0.3481$; *SNCA* uniform correlations: $r_{25\%} = 0.5653$, $r_{30\%} = 0.5780$, $r_{35\%} = 0.5767$, $r_{40\%} = 0.5794$; *TMEM175* uniform correlations: $r_{25\%} = 0.4483$, $r_{30\%} = 0.4051$, $r_{35\%} = 0.3866$, $r_{40\%} = 0.3758$, blue line in Fig. 6). Altogether, these results demonstrate that regional expression of *GBA* and *SNCA* shapes the spatial patterning of atrophy in addition to connectome topology and spatial embedding.

6.6 Structural and functional connectivity interact to drive disease spread

Finally, we tested whether neuronal activity or pre- and post-synaptic co-activation may facilitate α-synuclein propagation. Past neuroimaging studies have shown that cortical thinning in PD is predicted in part by functional connectivity to affected subcortical regions, and that regions that exhibit stronger functional connectivity with the substantia nigra tend to exhibit greater atrophy (Zeighami et al., 2015; Yau et al., 2018). Secretion of α-synuclein by neurons has been shown to be activity dependent (Paillusson et al., 2013). Spread of α-synuclein through multiple anatomical pathways may be biased by synchronous activity between the pre- and post-synaptic cells, such that the agents are more likely to move towards regions with higher functional connectivity to a seed region.

To address this question, we integrated resting-state fMRI functional connectivity into the model. We introduce a term $\mathrm{e}^{k\times \mathrm{fc}_{(i,j)}}$ to rescale the probability of moving from region i to region j previously defined by the connection strength of edge (i,j) while keeping the sum of the probabilities equal to 1 to preserve the multinomial distribution (see *Appendices*). As k is increased, the influence of functional connectivity is greater: stronger co-activation patterns play a more influential role in modulating the motion of the agents on structural connections. For structural edges with relatively small corresponding functional connectivity values, larger k may decrease those edges' contributions to favour propagation through edges with greater functional connectivity. All negative-valued and non-significant functional connections were set to zero.

We varied k from 0 (no influence of functional connectivity) to 5 and derived the corresponding peak values of model fit using the same four structural connectome densities as before (Fig. 7, red line). Model fit was improved by progressively increasing the importance of functional connectivity, but only up to a point ($k_{25\%} = 1, k_{30\%} = 2.5, k_{35\%} = 2.5, k_{40\%} = 2.5$). Beyond this point, the influence of functional connectivity dominates the agents' mobility pattern resulting in diminished model fit. The results were consistent across network densities. These results provide evidence for the notion that while α-synuclein propagation and resultant brain atrophy patterns occur via anatomical connections, they may also be biased by neuronal activity.

An alternative explanation is that inclusion of functional connectivity simply leads to over-fitting the model. To test this possibility, we investigated if the same improvement in model fit could be observed if α-synuclein spread is biased by randomized functional connectivity patterns. We generated "null" functional connectivity matrices by randomly re-assigning the parcellated rs-fMRI time series into the 42 left hemisphere regions (Fig. 7, grey bar). We note two important results. First, atrophy patterns based on real functional connectivity consistently yield significantly higher model fit than atrophy patterns based on null functional connectivity. Second, model fits based on null functional connectivity do not have the same peaked shape as observed when using real functional connectivity. This further support the conclusion that atrophy patterns observed in PD patients depend on both the structural and functional architecture of the brain.

7 Methods

7.1 Human brain parcellation

We used a brain parcellation generated by atlas-based segmentation (Cammoun et al., 2012). 68 cortical parcels were defined using curvature-based information (Desikan et al., 2006), which is available in FreeSurfer (http://surfer.nmr.mgh.harvard.edu). Subcortical parcels, including thalamus, caudate, putamen, pallidum, accumbens, amygdala, and hippocamppus, were extracted using the same software from a whole brain segmentation (Fischl et al., 2002). Finally, substantia nigra was added to the atlas using the location provided in the ATAG atlas (https://www.nitrc.org/projects/atag) (Keuken et al., 2014). Only the left hemisphere was used in this model, resulting in a total of 42 regions for the subsequent analysis. We used only the left hemisphere to simulate the propagation model because it is difficult to accurately determine interhemispheric connections using tractography (Jbabdi and Johansen-Berg, 2011). Moreover, regional gene expression was mostly available only for the left hemisphere (see *Regional gene expression*).

7.2 PPMI patient data and image processing

PPMI is an open-access comprehensive observational clinical study (Marek et al., 2011), longitudinally collecting multimodal imaging data, biological samples and clinical and behavioural assessments in a cohort of PD patients. 3T high-resolution T1-weighted MRI scans of 355 subjects (237 PD patients, 118 age-matched healthy controls) were obtained from the initial visit of PPMI to assess group-level regional atrophy using Deformation-Based Morphometry (DBM) (Zeighami et al., 2015), a method to detect local changes in tissue density.

After denoising (Coupé et al., 2008), inhomogeneity correction (Sled et al., 1998), and linear intensity scaling, individual MRI images are registered non-linearly to the MNI152-2009c template (Collins and Evans, 1997), yielding the corresponding transformation fields to be inverted into deformation maps in MNI space. Instead of directly using the displacement value $U(x) = (u_1(x), u_2(x), u_3(x))$ of voxel x at coordinates (x_1, x_2, x_3), we calculate the derivative of the displacement in each direction and take the determinant of the jacobian matrix J minus 1, namely, $|J| - 1$, as the value of deformation at x, which reflects local volume change.

$$J = \frac{\partial U}{\partial x} = \begin{pmatrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \frac{\partial u_1}{\partial x_3} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \frac{\partial u_2}{\partial x_3} \\ \frac{\partial u_3}{\partial x_1} & \frac{\partial u_3}{\partial x_2} & \frac{\partial u_3}{\partial x_3} \end{pmatrix} \tag{1}$$

These values constitute a 3D deformation map for each subject, on which an un-paired t test is conducted to derive the statistical difference (t-score) between the PD patients and the healthy controls at each voxel as a measure of local atrophy.

7.3 Regional gene expression

Regional gene expression levels were derived from the six post-mortem brains included in the Allen Human Brain Atlas (AHBA) (Hawrylycz et al., 2012), a multimodal atlas of the anatomy and microarray-based gene expression of the human brain. Individuals who donated their brains had no history of psychiatric or neurological disorders. Since four of the brains have data from the left hemisphere only, we only modeled the left hemisphere in our study, selecting a total of 3021 samples of *GBA* (probe ID: 1025373, 1025374) and *SNCA* (probe ID: 1020182, 1010655) in left hemisphere regions. Cortical samples were volumetrically mapped to the 34 cortical regions of our parcellation according to their corrected MNI coordinates (https://github.com/chrisfilo/alleninf) (Gorgolewski et al., 2014), also including samples that are within 1mm of the nearest gray matter coordinates assigned to any region. Subcortical samples were assigned to one of the 8 subcortical regions as specified by the structure names provided in the AHBA, due to imperfect registration of the post-mortem brains onto MNI space. For each probe, all samples that fell in the same anatomical region were averaged and then normalized across all 42 left hemisphere regions, generating transcription maps of each individual probe. These probe maps were next averaged according to the gene classification and normalized again across the regions, yielding the spatial expression profiles for *SNCA* and *GBA* respectively, represented as 42×1 vectors (Fig. S6).

7.4 Diffusion weighted image processing and structural connectivity

A total of 1027 subjects' preprocessed diffusion MRI data with the corresponding T1 images was obtained from the Human Connectome Project (2017 Q4, 1200-subject release) to construct an average macroscopic structural connectivity map of the healthy brain. With a multishell scheme of b values 1000, 2000, 3000 s/mm^2 and the number of diffusion sampling directions 90, 90, 90, the diffusion data were reconstructed in individual T1 spaces using generalized q-sampling imaing (GQI) (Yeh et al., 2010) with a diffusion sampling length ratio of 1.0, outputting at each voxel quantitative anisotropy (QA) and the Spin distribution function (SDF), a measurement of the density of diffusing water at different orientations (Yeh and Tseng, 2011).

Deterministic fiber tracking was conducted in native space using DSI studio (www.dsi-studio.lab
solver.org) (Yeh et al., 2013). The 42 left hemisphere regions in standard space were mapped non-linearly onto the individual T1 images using the FNIRT algorithm (https://fsl.fmrib.ox.ac.uk/) (Jenkinson et al., 2012) with a warp resolution of 8mm, 8mm, 8mm. The 34 cortical regions were dilated toward the grey-white matter interface by 1mm. The QA threshold was set to 0.6∗Otsu's threshold, which maximizes the variance between background and foreground

pixels. To compensate for volume-size introduced biases, deterministic tractography was performed for each region (taken as the seed region) separately. With an angular cutoff of 55, step size of 0.5mm, minimum length of 20mm, and maximum length of 400mm, 100,000 streamlines were reconstructed for each seed region. Connection strength between the seed region and the target region was set to be the density of streamlines (streamline counts) normalized by the volume size (voxel counts) of the target region and the mean length of the streamlines. The goal of this normalization is to correct for the bias toward large regions and long fibers inherent in the fiber tracking algorithms. The procedure was repeated for each region (as the tractography seed region), resulting in 42 connection profiles (42 1×42 vectors). Each connection profile consists of the connection strengths between the seed region and all other brain regions with self-connection set to zero. These connection profiles were finally concatenated to generate a 42×42 structural connectivity matrix per subject. Varying numbers of most commonly occurring edges were selected and averaged across the individual structural connectivity matrices to construct the group structural connectivity matrix with binary density ranging from 25% to 45%. These group-level matrices were finally symmetrized to represent (un-directed) brain networks. Likewise, we also constructed a group-level distance matrix in which elements denote mean euclidean length of the corresponding streamlines, which were used to model the mobility pattern of agents in the edges.

7.5 S-I-R agent-based model

The S-I-R agent-based model includes five modules:

- (a) production of normal α-synuclein
- (b) clearance of normal and misfolded α-synuclein
- (c) misfolding of normal α-synuclein (infection transmission)
- (d) propagation of normal and misfolded α-synuclein

- (e) accrual of neuronal death (atrophy).

It assumes that α-synuclein molecules are independent agents with mobility patterns and lifespans characterized by the connectome's architecture, neuronal activity, and regional gene expression. The normal α-synuclein agents, synthesized continuously under the modulation of regional *SNCA* expression, are susceptible to the misfolding process when they come in contact with a misfolded agent. Once infected, they adopt the misfolded form and join the infected population. Both normal and infected agents may spread via fiber tracts towards connected regions. The degradation rate of both agents is modulated by *GBA* expression, which codes for the lysosomal enzyme glucocerebrosidase (Sidransky and Lopez, 2012).

Production of normal α-synuclein. In each voxel of region i, a new normal agent may get synthesized per unit time with probability α_i, i.e., the synthesis rate in region i. α_i is chosen as $\Phi_{0,1}(SNCA\text{expression}_i)$ where $\Phi_{0,1}(\cdot)$ is the standard normal cumulative distribution function. Hence a higher expression score entails a higher α-synuclein synthesis rate. The increment of normal agents in region i is $\alpha_i S_i \Delta t$ after a total time Δt, where S_i is the size (voxel count) of region i. Δt was set to 0.01.

Clearance of normal and misfolded α-synuclein. Agents in region i, either normal or misfolded, may get cleared per unit time with probability β_i, the clearance rate in region i. As for synthesis rate, β_i is set to $\Phi_{0,1}(GBA\text{expression}_i)$. Considering that the probabililty that an agent is still active after a total time Δt is given by $\lim_{\delta\tau \to 0}(1-\beta\delta\tau)^{\Delta t/\delta\tau} = \mathrm{e}^{-\beta\Delta t}$, the cleared proportion within time step Δt is $1-\mathrm{e}^{-\beta\Delta t}$.

Misfolding of normal α-synuclein (infection transmission). The normal agents that survive clearance may become infected with probability $\gamma_i = 1-\mathrm{e}^{M_i \ln(1-\gamma_i^0)}$ in region i, where M_i is the population of misfolded agents and γ_i^0 is the baseline transmission rate which measures the likelihood that a single misfolded agent can transmit the infection to other susceptible agents. γ_i denotes the probability of getting infected by at least one of the M_i misfolded agents in region i per unit time. The baseline transmission rate γ_i^0 in region i is set to the reciprocal of region size, $1/S_i$. Analogous to the clearance module, the probability that a normal agent is uninfected after a total time Δt is given by $\lim_{\delta\tau\to 0}(1 - \gamma_i^0 \delta\tau)^{M_i \Delta t/\delta\tau} = \mathrm{e}^{-\gamma_i^0 M_i \Delta t}$, thus the proportion of normal agents that undergo misfolding within Δt is $1 - \mathrm{e}^{-\gamma_i^0 M_i \Delta t}$.

Therefore, in determining the baseline regional density of normal α-synuclein, we increment the population of normal agents N_i with:

$$\Delta N_i = \alpha_i S_i \Delta t - (1 - \mathrm{e}^{-\beta_i \Delta t}) N_i \tag{2}$$

After the system reaches the stable point (error tolerance $\epsilon < 10^{-7}$), we initiate the pathogenic spread and update the population of normal (N) and misfolded (M) agents with:

$$\begin{aligned} \Delta N_i =& \alpha_i S_i \Delta t - (1 - \mathrm{e}^{-\beta_i \Delta t}) N_i - (\mathrm{e}^{-\beta_i \Delta t})(1 - \mathrm{e}^{-\gamma_i^0 M_i \Delta t}) N_i & (3) \\ \Delta M_i =& (\mathrm{e}^{-\beta_i \Delta t})(1 - \mathrm{e}^{-\gamma_i^0 M_i \Delta t}) N_i - (1 - \mathrm{e}^{-\beta_i \Delta t}) M_i & (4) \end{aligned}$$

The system has two fixed points, whose final positions will not be affected by the initial conditions of (N_i, M_i), including the choice of seed region and seeded misfolded agents (see *Appendices*). Note that normal and misfolded agents are equivalent to susceptible and infected agents.

Propagation of normal and misfolded α-synuclein. Agents in region i may remain in region i or enter the edges according to a multinomial distribution per unit time with probabilities:

$$\begin{aligned} P_{\text{region } i \to \text{region } i} &= \rho_i & (5) \\ P_{\text{region } i \to \text{edge } (i,j)} &= (1 - \rho_i) \frac{w_{ij}}{\sum_j w_{ij}} & (6) \end{aligned}$$

where w_{ij} is the connection strength of edge (i, j) (fiber tracts density between region i and j). The probability of remaining in the current region i, ρ_i, was set to 0.5 for all i (see Fig. S7(a) for other choices of ρ_i; we note that the model fit is robust to variations in ρ_i).

Analogously, the agents in edge (i, j) may exit the edge or remain in the same edge per unit time with binary probabilities:

$$P_{\text{edge } (i,j) \to \text{region } j} = \frac{1}{l_{ij}} \tag{7}$$

$$P_{\text{edge } (i,j) \to \text{edge } (i,j)} = 1 - \frac{1}{l_{ij}} \tag{8}$$

where l_{ij} is the length of edge (i, j) (the mean length of the fiber tracts between region i and region j). In the absence of definitive molecular evidence of different spreading rates for normal and misfolded α-synuclein, we do not assume different exit and propagation dynamics for the two types of agents. We use $N_{(i,j)}, M_{(i,j)}$ to denote the normal/misfolded population in edge (i, j). After a total time Δt, the increments of N_i, M_i in region i are:

$$\Delta N_i = \sum_j \frac{1}{l_{ji}} N_{(j,i)} \Delta t - (1 - \rho_i) N_i \Delta t \tag{9}$$

$$\Delta M_i = \sum_j \frac{1}{l_{ji}} M_{(j,i)} \Delta t - (1 - \rho_i) M_i \Delta t \tag{10}$$

Likewise,

$$\Delta N_{(i,j)} = (1 - \rho_i) \frac{w_{ij}}{\sum_j w_{ij}} N_i \Delta t - \frac{1}{l_{ij}} N_{(i,j)} \Delta t \tag{11}$$

$$\Delta M_{(i,j)} = (1 - \rho_i) \frac{w_{ij}}{\sum_j w_{ij}} M_i \Delta t - \frac{1}{l_{ij}} M_{(i,j)} \Delta t \tag{12}$$

We adopt an asynchronous implementation in which the propagation of normal and misfolded agents is operated before the synthesis, clearance and infection at each Δt. We have also tried other implementations, including propagation after synthesis/clearance/infection at each Δt and synchronous implementation, and found the differences are negligible, suggesting that our results are independent of the modules' update order. Note that although the agent-based model can also be viewed in a stochastic framework (i.e. individual agents alter their states stochastically and the total number of agents at any time is discrete valued), we conducted the simulations in a deterministic way (i.e. using the mean values for each subpopulation of agents in a region, which can take on non-integer values), which preserves the dynamics of disease spreading as the population of protein agents is sufficiently large.

Accrual of neuronal death (atrophy). We model neuronal death as the result of two processes: direct toxicity from accumulation of native misfolded α-synuclein and deafferentation (reduction in neuornal inputs) from neuronal death in neighbouring (connected) regions. The atrophy accrual at t within Δt in region i is given by the sum of the two processes:

$$\Delta L_i(t) = k_1(1 - \mathrm{e}^{-r_i(t)\Delta t}) + k_2 \sum_j \frac{w_{ji}}{\sum_j w_{ji}}(1 - \mathrm{e}^{-r_j(t-1)\Delta t}) \tag{13}$$

where $r_i(t)$ is the proportion of misfolded agents in region i at time t, and $1 - \mathrm{e}^{-r_i(t)\Delta t}$ quantifies the increment of atrophy caused by accumulation of native misfolded α-synuclein aggregates within Δt at time t. The second term $1 - \mathrm{e}^{-r_j(t-1)\Delta t}$, weighted by $w_{ji}/\sum_j w_{ji}$ and summed up across j, accounts for the increment of atrophy induced by deafferentation from neighbouring regions within Δt at $t-1$. k_1, k_2 are the weights of the two terms with $k_1 + k_2 = 1$. We set $k_1 = k_2 = 0.5$ such that native α-synuclein accumulation and the deafferentation have equal importance in modelling the total atrophy growth (see Fig. S7(b) for other choices of k_1, k_2; we note that the model fit is consistent across k_1/k_2 ranging from 0.1 to 10).

7.6 Integration of functional connectivity

We used resting-state functional MRI (fMRI) scans from the Human Connectome Project (2015, S500 release) to construct the functional connectivity maps. Both left-right (LR) and right-left (RL) phase encoding direction data were used. Based on the minimally preprocessed resting-state fMRI data, further processing steps were performed, including: 1) nuisance signal regression (including white matter, cerebrospinal fluid, global signal, and six motion parameters); 2) bandpass temporal filtering (0.01 Hz f 0.08 Hz); 3) spatial smoothing using a 4mm FWHM Gaussian kernel. After quality control, 494 subjects were finally included.

We then extracted the mean time course in each of the 42 regions and computed the pairwise Pearson's correlation coefficients to derive individual functional connectivity matrices. Normalized by Fisher's z transform, the functional connectivity matrices were averaged across subjects and converted back to correlations using inverse Fisher transform to generate the group functional connectivity matrix. All negative correlations in the resultant functional connectivity matrix were set to zero, having no influence on the agents' mobility pattern.

Integration of functional connectivity into the model should bias mobility of the agents towards region pairs showing greater co-activation patterns. Agents thus have a higher chance of entering the edges that connect regions having stronger synchronous neuronal activity. More specifically, the weights w_{ij} (connection strength of structural connectivity) in equation (6) were scaled by $\mathrm{e}^{k\times \mathrm{fc}_{(i,j)}}$, where $\mathrm{fc}_{(i,j)}$ is the functional connectivity between region i and region j. Therefore the probability that agents move from region i to edge (i,j) per unit time is determined by

$$P_{\text{region } i\to\text{edge } (i,j)} = (1-\rho_i)\frac{\mathrm{e}^{k\times \mathrm{fc}_{(i,j)}} w_{ij}}{\sum_j \mathrm{e}^{k\times \mathrm{fc}_{(i,j)}} w_{ij}} \tag{14}$$

Note that increasing k makes the influence of functional connectivity more differentiated across the edges: the stronger functional connectivity values will be enhanced while the weaker ones may be suppressed.

8 Discussion

We developed a networked S-I-R agent-based model of neurodegenerative disease consisting of normal and misfolded proteins. Taking PD as an example, we integrated multimodal neuroimaging and gene expression data to simulate the propagation of misfolded α-synuclein on the healthy connectome. The S-I-R agent-based model incorporates pathogenic spread (dominated by the connectome) and selective vulnerability (modulated here by gene expression) under the same computational framework. The dynamic model replicated the spatial pattern of measured brain atrophy in PD patients and had greater predictive power than any of the constituent features (i.e. network metrics or gene expression) on their own. Our results demonstrate that connectome topology and geometry, local gene expression, and functional co-activation jointly shape disease progression, as systematic disruption of each of these elements significantly degraded model performance. This model yields insights into the mechanism of PD, providing support for the propagating proteinopathy theory, and can be readily adapted to other neurodegenerative diseases.

Network-based epidemic models can be classified into two main categories: (1) diffusion models, in which nodes of the graph adopt different states (e.g. concentration of disease), and the infection propagates via diffusion; and (2) agent-based models, in which agents (in this case, proteins) adopt different states (S-I-R) and interact with each other according to simple rules. The S-I-R model used here has reduced complexity compared to the pure agent-based model but preserves individual stochasticity and regional heterogeneity. It uses a compartmental model within (brain) regions, but allows agents to travel along the connectome (Balcan et al., 2010). This allows us to map the interaction between network architecture and regional susceptibility and transmissibility. Solving our S-I-R agent-based model numerically yielded two fixed (or stable) points of the process after seeding of the infection: rapid extinction or epidemic spread. If the system is attracted to extinction, misfolded proteins will eventually be cleared. If the system is attracted to the fixed point, this represents an outbreak: misfolded proteins will still not accumulate boundlessly, but will finally achieve a stable concentration and co-exist with normal proteins. These results are consistent with recent experimental evidence in rodents where injected misfolded α-synuclein fibrils either died out at specific injection cites (extinction), or grew but ceased to propagate at a certain stage (Rey et al., 2018), suggesting the existence of an equilibrium. These different outcomes (extinction vs outbreak) might perhaps represent normal aging versus progressive neurodegeneration, or mild versus malignant PD (Fereshtehnejad et al., 2017).

A priori there are several topological properties of brain networks that favour disease spread. Brains are canonical examples of small-world networks (Watts and Strogatz, 1998), where path lengths between any two nodes are relatively short, a feature that facilitates disease outbreaks (Moore and Newman, 2000). Brain networks display prominent community structure: the tendency to cluster into highly interconnected communities or modules (Hilgetag and Kaiser, 2004; Sporns and Betzel, 2016). Perhaps counter-intuitively, community structure may also potentiate global disease spread by enhancing local, intra-community infection (Nematzadeh et al., 2014). Finally, the presence of high degree nodes (hubs) that are highly interconnected with each other favours disease propagation (van den Heuvel et al., 2012). Hubs and are expected to have faster arrival times, and greater accumulation of infected agents, making them especially vulnerable to attack. Indeed, hubs manifest greater structural abnormalities in a host of neurodegenerative diseases (Crossley et al., 2014). We showed that disruptions of the network's architecture reduce model fit, providing evidence that the emergent dynamics of synucleinopathy depend on network connectivity (topology) and geometry (spatial embedding).

While we did find that network metrics predict brain atrophy, the full S-I-R agent-based model provided a better fit to the empirical data than these metrics on their own (Fig. 3). Spatial proximity among regions and local differences in synthesis/clearance (see below) both impose constraints on the spreading process, amplifying or damping the rate of infection. As a result, atrophy patterns are shaped by, but ultimately transcend, the underlying connection patterns. The present model correctly predicts that the regions most vulnerable to atrophy are not simply those that participate in the greatest number of connections or those that are a few steps away from other infected regions. A similar phenomenon was recently described in schizophrenia: vulnerability of prefrontal "hub" regions to grey matter atrophy may be a function of both connectivity and spatial location (Gollo et al., 2018).

More specifically, the agent-based model also allowed us to test two competing theories of PD pathogenesis: protein propagation versus regional vulnerability (Brundin and Melki, 2017; Surmeier et al., 2017). Here we chose to model regional vulnerability by incorporating estimated local α-synuclein concentration, known to facilitate seeding (Volpicelli-Daley et al., 2011) and increase neuronal vulnerability in animal models (Luna et al., 2018). We used regional expression of *GBA* and *SNCA* as estimates of α-synuclein clearance and synthesis rates to derive the concentration of endogenous α-synuclein. We showed that incorporating this information into the model improved the correlation with empirical atrophy in PD patients; moreover, spatial permutation of gene expression degraded the fit. Thus, our findings support the theory that the dynamics of disease progression arise from an interplay between regional vulnerability and network-wide propagation.

Our results provide converging evidence for the involvement of *GBA* and *SNCA* in PD pathology previously indicated in animal and cellular studies (Alcalay et al., 2015). Mutations in *GBA* are the most common genetic risk factor for PD (Aharon-Peretz et al., 2004; Sidransky and Lopez, 2012); mutations and multiplications of *SNCA* have been implicated in driving the severity of the pathology (Singleton et al., 2003; Chartier-Harlin et al., 2004; Ibanez et al., 2004). It is worth noting that simple spatial correlation measures alone failed to uncover the effects of *GBA* or *SNCA* regional expression on the empirical atrophy pattern; the gene expression effects only emerged from the full agent-based propagating model, which therefore provides a new way to identify gene-disease associations in the central nervous system. New genes can easily be incorporated to adapt the model to other neurodegenerative diseases.

The benefit of augmenting network structure with protein spreading dynamics is exemplified by the identification of the substantia nigra as the likeliest disease epicenter. Mirroring the *Reproduction number* R_0 (Newman, 2010), which marks the transition between disease extinction and outbreak in conventional epidemic models, we estimated spread threshold for our S-I-R agent-based model. This represents the minimum number of infectious agents that need to be introduced in any area that will lead to a major outbreak. In our model, the substantia nigra has the lowest spread threshold, identifying it as a likely disease epicenter. This is consistent with the Braak hypothesis, where the substantia nigra is one of the earliest affected sites in the central nervous system. Note that we could not include structures in the pons and medulla, thought to be affected even earlier in PD (Braak et al., 2003, 2004), due to difficulty in imaging either brain atrophy or white matter tracts in the brainstem.

It is also known that α-synuclein is secreted in an activity dependent manner (Paillusson et al., 2013). We therefore tested the influence of resting state-fMRI derived measures of functional connectivity on protein mobility. As a measure of synchronous neuronal activity in pre- and post-synaptic regions, functional connectivity will bias the proteins into regions showing greater co-activation. Once again we found that this addition significantly improved the model fit. Thus, functional co-activation also shapes the pattern of disease propagation, explaining why atrophy patterns in neurodegenerative diseases tend to resemble intrinsic functional networks (Seeley et al., 2009; Zeighami et al., 2015).

We took advantage of several useful features of agent-based models to provide an understanding of factors involved in disease propagation. Others have applied more traditional diffusion models to Alzheimer's Disease (Raj et al., 2012) and to neurodegeneration more generally (Weickenmeier et al., 2018b); however the agent-based model used here affords us the possibility of testing different mechanisms of disease, likelihood of outbreak, effect of emergent properties (such as the effect of regional neuronal death on subsequent disease propagation) and, eventually, therapeutic interventions. Note that we divided the population into compartments in which the agents share the same characteristics, making the spreading dynamics more tractable and computationally efficient. They can be easily tailored to accommodate a full agent-based setting by introducing more fine-grained rules. For example, the transmission rate $\gamma_i = 1 - \mathrm{e}^{M \ln(1-\gamma_i^0)}$ can be extrapolated as $\gamma_i = 1 - \mathrm{e}^{\sum_{k=1}^{M} \ln(1-\gamma_{i,k}^0)}$ to model individually differentiated transmission rates $\gamma_{i,k}^0$ in region i (see *Appendices*).

Although the S-I-R agent-based model provided a good fit to observed neurodegeneration, there are several caveats and limitations in the present study. First, regional variations in vulnerability to the toxicity of misfolded α-synuclein apart from the effects of α-synuclein concentration were not accounted for. Moreover, neuronal loss was homogeneously modelled as a simple linear combination of cell death due to native α-synuclein accumulation and deafferentation, which may not reflect reality. It is possible that regions respond differently to the toxicity of α-synuclein aggregates, and this can easily be incorporated into the model by introducing new factors, such as genes that control resilience to energetic stress for example (Michel et al., 2016). Moreover, cell death may slow the propagation of misfolded α-synuclein and accrual of atrophy, especially in more affected regions. Although we did not take this effect into account here, it can easily be incorporated into the model using agent-based rules.

Second, the white matter network may not represent the exact physical routes of spread. It is possible that α-synuclein spread occurs only between specific cell types, or in one direction, while, in our model, the agents spread bi-directionally along the fiber tracts. The outlier region (accumbens, Fig. 2(b)) that remarkably impedes model fit serves as an example. Nucleus accumbens is one of the least atrophied regions in the dataset used here, whereas it exhibits high atrophy in the model. One possible reason for this disagreement is that we did not include the different subsections of the substantia nigra and their projections in the structural connectome used for the model. While we seeded the entire substantia nigra, it is known that the medial portion, which projects to the accumbens (Zhang et al., 2017), is less affected in PD than the lateral substantia nigra, which projects to dorsal striatum (Braak et al., 2003, 2004).

Finally, we focused on only two genes in modelling synucleinopathy, while many other genes such as *LRKK2* and *MAPT*, and proteins such as dopamine or *tau*, may also influence or interact with synucleinopathy propagation. Using a small subset of genes avoids high model complexity and allowed modelling the proteinopathy in a parameter-free setting. However, the parameter-free setting introduces another caveat: the model converts gene expression scores and fiber density into probabilities without scaling their relative magnitude, while the actual rate of synthesis/clearance and protein spreading may not be at the same scale. In the future this can be solved by introducing scaling parameters into the model, training the model on individual connectome and genetic data to find the optimal solutions, and including multiple timepoints from patient data.

One of the future directions is to customize the model with individual anatomical, functional, genetic or clinical data to increase its ability to predict disease trajectory and to identify factors that promote resistance to disease spread. Moreover, this model can hopefully help test new preventative procedures. Introducing medications may change the parameters of the dynamical system; for example, increasing *GBA* activity to elevate the clearance rate would make the stable point for extinction more robust to small perturbations (e.g., the invasion of misfolded α-synuclein fibrils).

9 Conclusions

First, we have revealed how proteinopathy propagation on brain network and local differences in vulnerability interactively drive disease spread. The connectome architecture determines the physical routes of spreading, capturing the effect of "pathogenic spread" via neuronal connections. We found in null networks in which the architecture of the connectome (either spatial embedding or topology) is disrupted, the model fit with the real atrophy pattern is remarkably degraded, indicating proteinopathy propagation on brain network serves as a major driving force of disease progression. Moreover, the model with uniform transcription profiles underperformed compared to those using real gene expressions, suggesting the disease spread cannot be solely explained by "pathogenic spread" but also depends on local differences in vulnerability modulated by genes.

Second, we have identified the factors that shape the disease spread. Neuronal connections (measured by structural connectivity) dictates the spreading of pathological proteins, while neural activity (measured by functional connectivity) bias the proteins toward regions showing stronger co-activation pattern with the previous regions, on top of the physical routes determined by neuronal connections. Additionally, *SNCA*, *GBA* and *TMEM175* modulating regional variations in vulnerability by influencing local protein synthesis, seeding and clearance.

Last, we have shown S-I-R agent-based model as a powerful tool to recover disease trajectory of neurodegeneration. Although entirely governed by information from the healthy brain, the model can recapitulate the distribution of neuronal loss identified from an independent T1 dataset of PD patients and healthy controls, and has more predicative power than its constituent factors *per se*. The agent-based SIR model is a valid approach to integrate the multimodal disease-relevant factors to describe longitudinal disease progression. By taking proteins as individual agents and modelling their behaviour using macroscale measures (connectome, neural activity, and regional gene expressions), it bridges the microscopic properties of protein interaction and the macroscopic consequence of disease spreading. Compared to conventional network diffusion models, the agent-based S-I-R model is more biologically meaningful (i.e. taking into account the reseeding process of pathological proteins rather than simply model the spreading driven by concentration gradients). More importantly, it allows easy testing the effects of therapeutic procedures in intervening disease spread, for example, boosting *GBA* expression to increase clearance rate thereby prevents the outbreak of infectious proteins.

10 Bibliography

Aharon-Peretz, J., Rosenbaum, H., and Gershoni-Baruch, R. (2004). Mutations in the glucocerebrosidase gene and parkinson's disease in ashkenazi jews. *New England Journal of Medicine*, 351(19):1972–1977.

Alcalay, R. N., Levy, O. A., Waters, C. H., Fahn, S., Ford, B., Kuo, S.-H., Mazzoni, P., Pauciulo, M. W., Nichols, W. C., Gan-Or, Z., et al. (2015). Glucocerebrosidase activity in parkinson's disease with and without gba mutations. *Brain*, 138(9):2648–2658.

Balcan, D., Gonçalves, B., Hu, H., Ramasco, J. J., Colizza, V., and Vespignani, A. (2010). Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *Journal of computational science*, 1(3):132–145.

Braak, H., Del Tredici, K., Rüb, U., De Vos, R. A., Steur, E. N. J., and Braak, E. (2003). Staging of brain pathology related to sporadic parkinson's disease. *Neurobiology of aging*, 24(2):197–211.

Braak, H., Ghebremedhin, E., Rüb, U., Bratzke, H., and Del Tredici, K. (2004). Stages in the development of parkinson's disease-related pathology. *Cell and tissue research*, 318(1):121–134.

Brettschneider, J., Del Tredici, K., Lee, V. M.-Y., and Trojanowski, J. Q. (2015). Spreading of pathology in neurodegenerative diseases: a focus on human studies. *Nature Reviews Neuroscience*, 16(2):109.

Brundin, P. and Kordower, J. H. (2012). Neuropathology in transplants in parkinson's disease: implications for disease pathogenesis and the future of cell therapy. In *Progress in brain research*, volume 200, pages 221–241. Elsevier.

Brundin, P. and Melki, R. (2017). Prying into the prion hypothesis for parkinson's disease. *Journal of Neuroscience*, 37(41):9808–9818.

Cammoun, L., Gigandet, X., Meskaldji, D., Thiran, J. P., Sporns, O., Do, K. Q., Maeder, P., Meuli, R., and Hagmann, P. (2012). Mapping the human connectome at multiple scales with diffusion spectrum mri. *Journal of neuroscience methods*, 203(2):386–397.

Chartier-Harlin, M.-C., Kachergus, J., Roumier, C., Mouroux, V., Douay, X., Lincoln, S., Levecque, C., Larvor, L., Andrieux, J., Hulihan, M., et al. (2004). α-synuclein locus duplication as a cause of familial parkinson's disease. *The Lancet*, 364(9440):1167–1169.

Collins, D. L. and Evans, A. C. (1997). Animal: validation and applications of nonlinear registration-based segmentation. *International journal of pattern recognition and artificial intelligence*, 11(08):1271–1294.

Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., and Barillot, C. (2008). An optimized blockwise nonlocal means denoising filter for 3-d magnetic resonance images. *IEEE transactions on medical imaging*, 27(4):425–441.

Crossley, N. A., Mechelli, A., Scott, J., Carletti, F., Fox, P. T., McGuire, P., and Bullmore, E. T. (2014). The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain*, 137(8):2382–2395.

Damier, P., Hirsch, E., Agid, Y., and Graybiel, A. (1999). The substantia nigra of the human brain: Ii. patterns of loss of dopamine-containing neurons in parkinson's disease. *Brain*, 122(8):1437–1448.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980.

Fereshtehnejad, S.-M., Zeighami, Y., Dagher, A., and Postuma, R. B. (2017). Clinical criteria for subtyping parkinson's disease: biomarkers and longitudinal progression. *Brain*, 140(7):1959–1976.

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355.

Frias-Martinez, E., Williamson, G., and Frias-Martinez, V. (2011). An agent-based model of epidemic spread using human mobility and social network information. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*, pages 57–64. IEEE.

Goedert, M., Spillantini, M. G., Del Tredici, K., and Braak, H. (2013). 100 years of lewy pathology. *Nature Reviews Neurology*, 9(1):13.

Gollo, L. L., Roberts, J. A., Cropley, V. L., Di Biase, M. A., Pantelis, C., Zalesky, A., and Breakspear, M. (2018). Fragility and volatility of structural hubs in the human connectome. *Nature neuroscience*, 21(8):1107.

Gorgolewski, K., Fox, A., Chang, L., Schäfer, A., Arélin, K., Burmann, I., Sacher, J., and Margulies, D. (2014). Tight fitting genes: Finding relations between statistical maps and gene expression patterns. *Organization for Human Brain Mapping. Hamburg, Germany.*

Guo, J. L. and Lee, V. M. (2014). Cell-to-cell transmission of pathogenic proteins in neurodegenerative diseases. *Nature medicine*, 20(2):130.

Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., Van De Lagemaat, L. N., Smith, K. A., Ebbert, A., Riley, Z. L., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, 489(7416):391.

Hilgetag, C. C. and Kaiser, M. (2004). Clustered organization of cortical connectivity. *Neuroinformatics*, 2(3):353–360.

Ibanez, P., Bonnet, A., Debarges, B., Lohmann, E., Tison, F., Agid, Y., Dürr, A., Brice, A., Pollak, P., Group, F. P. D. G. S., et al. (2004). Causal relation between α-synuclein locus duplication as a cause of familial parkinson's disease. *The Lancet*, 364(9440):1169–1171.

Jackson, W. S. (2014). Selective vulnerability to neurodegenerative disease: the curious case of prion protein. *Disease models & mechanisms*, 7(1):21–29.

Jbabdi, S. and Johansen-Berg, H. (2011). Tractography: where do we go from here? *Brain connectivity*, 1(3):169–183.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage*, 62(2):782–790.

Jinn, S., Drolet, R. E., Cramer, P. E., Wong, A. H.-K., Toolan, D. M., Gretzula, C. A., Voleti, B., Vassileva, G., Disa, J., Tadin-Strapps, M., et al. (2017). Tmem175 deficiency impairs lysosomal and mitochondrial function and increases α-synuclein aggregation. *Proceedings of the National Academy of Sciences*, 114(9):2389–2394.

Jucker, M. and Walker, L. C. (2013). Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature*, 501(7465):45.

Keuken, M. C., Bazin, P.-L., Crown, L., Hootsmans, J., Laufer, A., Müller-Axt, C., Sier, R., van der Putten, E., Schäfer, A., Turner, R., et al. (2014). Quantifying inter-individual anatomical variability in the subcortex using 7t structural mri. *NeuroImage*, 94:40–46.

Kordower, J. H., Chu, Y., Hauser, R. A., Freeman, T. B., and Olanow, C. W. (2008). Lewy body–like pathology in long-term embryonic nigral transplants in parkinson's disease. *Nature medicine*, 14(5):504.

Li, J.-Y., Englund, E., Holton, J. L., Soulet, D., Hagell, P., Lees, A. J., Lashley, T., Quinn, N. P., Rehncrona, S., Björklund, A., et al. (2008). Lewy bodies in grafted neurons in subjects with parkinson's disease suggest host-to-graft disease propagation. *Nature medicine*, 14(5):501.

Luk, K. C., Kehm, V., Carroll, J., Zhang, B., O'Brien, P., Trojanowski, J. Q., and Lee, V.

M.-Y. (2012). Pathological α-synuclein transmission initiates parkinson-like neurodegeneration in nontransgenic mice. *Science*, 338(6109):949–953.

Luna, E., Decker, S. C., Riddle, D. M., Caputo, A., Zhang, B., Cole, T., Caswell, C., Xie, S. X., Lee, V. M., and Luk, K. C. (2018). Differential α-synuclein expression contributes to selective vulnerability of hippocampal neuron subpopulations to fibril-induced toxicity. *Acta neuropathologica*, pages 1–21.

Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., et al. (2011). The parkinson progression marker initiative (ppmi). *Progress in neurobiology*, 95(4):629–635.

Maslov, S. and Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science*, 296(5569):910–913.

Masuda-Suzukake, M., Nonaka, T., Hosokawa, M., Oikawa, T., Arai, T., Akiyama, H., Mann, D. M., and Hasegawa, M. (2013). Prion-like spreading of pathological α-synuclein in brain. *Brain*, 136(4):1128–1138.

Mattson, M. P. and Magnus, T. (2006). Ageing and neuronal vulnerability. *Nature Reviews Neuroscience*, 7(4):278.

Michel, P. P., Hirsch, E. C., and Hunot, S. (2016). Understanding dopaminergic cell death pathways in parkinson disease. *Neuron*, 90(4):675–691.

Mišić, B., Betzel, R. F., Griffa, A., de Reus, M. A., He, Y., Zuo, X.-N., van den Heuvel, M. P., Hagmann, P., Sporns, O., and Zatorre, R. J. (2018). Network-based asymmetry of the human auditory system. *Cerebral Cortex*, 28(7):2655–2664.

Mišić, B., Betzel, R. F., Nematzadeh, A., Goni, J., Griffa, A., Hagmann, P., Flammini, A., Ahn, Y.-Y., and Sporns, O. (2015). Cooperative and competitive spreading dynamics on the human connectome. *Neuron*, 86(6):1518–1529.

Moore, C. and Newman, M. E. (2000). Epidemics and percolation in small-world networks. *Physical Review E*, 61(5):5678.

Mougenot, A.-L., Nicot, S., Bencsik, A., Morignat, E., Verchère, J., Lakhdar, L., Legastelois, S., and Baron, T. (2012). Prion-like acceleration of a synucleinopathy in a transgenic mouse model. *Neurobiology of aging*, 33(9):2225–2228.

Nematzadeh, A., Ferrara, E., Flammini, A., and Ahn, Y.-Y. (2014). Optimal network modularity for information diffusion. *Physical review letters*, 113(8):088701.

Newman, M. (2010). *Networks: an introduction*. Oxford university press.

Paillusson, S., Clairembault, T., Biraud, M., Neunlist, M., and Derkinderen, P. (2013). Activity-dependent secretion of alpha-synuclein by enteric neurons. *Journal of neurochemistry*, 125(4):512–517.

Pandya, S., Zeighami, Y., Freeze, B., Dadar, M., Collins, D., Dagher, A., and Raj, A. (2019). Predictive model of spread of parkinson's pathology using network diffusion. *NeuroImage*.

Peelaerts, W., Bousset, L., Van der Perren, A., Moskalyuk, A., Pulizzi, R., Giugliano, M., Van Den Haute, C., Melki, R., and Baekelandt, V. (2015). α-synuclein strains cause distinct synucleinopathies after local and systemic administration. *Nature*, 522(7556):340.

Polymenidou, M. and Cleveland, D. W. (2012). Prion-like spread of protein aggregates in neurodegeneration. *Journal of Experimental Medicine*, 209(5):889–893.

Raj, A., Kuceyeski, A., and Weiner, M. (2012). A network diffusion model of disease progression in dementia. *Neuron*, 73(6):1204–1215.

Rey, N. L., George, S., Steiner, J. A., Madaj, Z., Luk, K. C., Trojanowski, J. Q., Lee, V. M.-Y., and Brundin, P. (2018). Spread of aggregates after olfactory bulb injection of α-synuclein fibrils is associated with early neuronal loss and is reduced long term. *Acta neuropathologica*, 135(1):65–83.

Rey, N. L., Steiner, J. A., Maroof, N., Luk, K. C., Madaj, Z., Trojanowski, J. Q., Lee, V. M.-Y., and Brundin, P. (2016). Widespread transneuronal propagation of α-synucleinopathy triggered in olfactory bulb mimics prodromal parkinson's disease. *Journal of Experimental Medicine*, pages jem–20160368.

Roberts, J. A., Perry, A., Lord, A. R., Roberts, G., Mitchell, P. B., Smith, R. E., Calamante, F., and Breakspear, M. (2016). The contribution of geometry to the human connectome. *Neuroimage*, 124:379–393.

Rubinov, M. and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069.

Saxena, S. and Caroni, P. (2011). Selective neuronal vulnerability in neurodegenerative diseases: from stressor thresholds to degeneration. *Neuron*, 71(1):35–48.

Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L., and Greicius, M. D. (2009). Neurodegenerative diseases target large-scale human brain networks. *Neuron*, 62(1):42–52.

Seguin, C., van den Heuvel, M. P., and Zalesky, A. (2018). Navigation of brain networks. *Proceedings of the National Academy of Sciences*, page 201801351.

Sidransky, E. and Lopez, G. (2012). The link between the gba gene and parkinsonism. *The Lancet Neurology*, 11(11):986–998.

Singleton, A., Farrer, M., Johnson, J., Singleton, A., Hague, S., Kachergus, J., Hulihan, M., Peuralinna, T., Dutra, A., Nussbaum, R., et al. (2003). α-synuclein locus triplication causes parkinson's disease. *Science*, 302(5646):841–841.

Sled, J. G., Zijdenbos, A. P., and Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE transactions on medical imaging*, 17(1):87–97.

Spillantini, M. G., Crowther, R. A., Jakes, R., Hasegawa, M., and Goedert, M. (1998). 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.

Sporns, O. and Betzel, R. F. (2016). Modular brain networks. *Annual review of psychology*, 67:613–640.

Surmeier, D. J., Obeso, J. A., and Halliday, G. M. (2017). Selective neuronal vulnerability in parkinson disease. *Nature Reviews Neuroscience*, 18(2):101.

van den Heuvel, M. P., Kahn, R. S., Goñi, J., and Sporns, O. (2012). High-cost, high-capacity backbone for global brain communication. *Proceedings of the National Academy of Sciences*, 109(28):11372–11377.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., Consortium, W.-M. H., et al. (2013). The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79.

Volpicelli-Daley, L. A., Luk, K. C., Patel, T. P., Tanik, S. A., Riddle, D. M., Stieber, A., Meaney, D. F., Trojanowski, J. Q., and Lee, V. M.-Y. (2011). Exogenous α-synuclein fibrils induce lewy body pathology leading to synaptic dysfunction and neuron death. *Neuron*, 72(1):57–71.

Walsh, D. M. and Selkoe, D. J. (2016). A critical appraisal of the pathogenic protein spread hypothesis of neurodegeneration. *Nature Reviews Neuroscience*, 17(4):251.

Watts, D. J. (2002). A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, 393(6684):440.

Weickenmeier, J., Kuhl, E., and Goriely, A. (2018a). Multiphysics of prionlike diseases: Progression and atrophy. *Physical review letters*, 121(15):158101.

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

Yau, Y., Zeighami, Y., Baker, T., Larcher, K., Vainik, U., Dadar, M., Fonov, V., Hagmann, P., Griffa, A., Mišić, B., et al. (2018). Network connectivity determines cortical thinning in early parkinson's disease progression. *Nature communications*, 9(1):12.

Yeh, F.-C. and Tseng, W.-Y. I. (2011). Ntu-90: a high angular resolution brain atlas constructed by q-space diffeomorphic reconstruction. *Neuroimage*, 58(1):91–99.

Yeh, F.-C., Verstynen, T. D., Wang, Y., Fernández-Miranda, J. C., and Tseng, W.-Y. I. (2013). Deterministic diffusion fiber tracking improved by quantitative anisotropy. *PloS one*, 8(11):e80713.

Yeh, F.-C., Wedeen, V. J., and Tseng, W.-Y. I. (2010). Generalized q-sampling imaging. *IEEE transactions on medical imaging*, 29(9):1626–1635.

Zeighami, Y., Ulla, M., Iturria-Medina, Y., Dadar, M., Zhang, Y., Larcher, K. M.-H., Fonov, V., Evans, A. C., Collins, D. L., and Dagher, A. (2015). Network structure of brain atrophy in de novo parkinson's disease. *Elife*, 4.

Zhang, Y., Larcher, K. M.-H., Misic, B., and Dagher, A. (2017). Anatomical and functional organization of the human substantia nigra and its connections. *Elife*, 6:e26653.

11 List of Figures

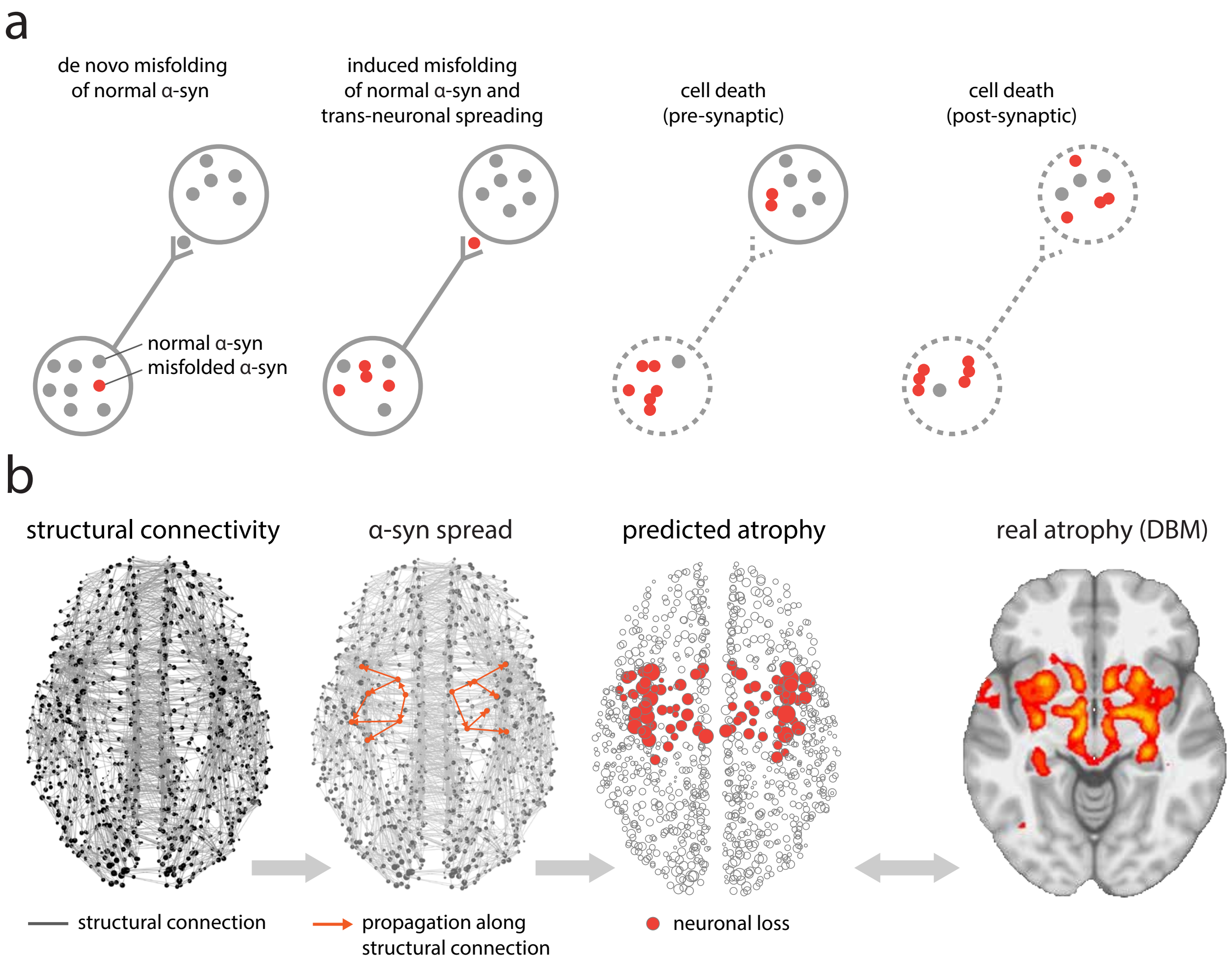

Figure 1: **Agent-based S-I-R model** — **(a)** Misfolded α-synuclein (red) may diffuse through synaptic connections into adjacent neurons, causing misfolding of normal α-synuclein (gray). Accumulation of misfolded α-synuclein induces neuronal loss. **(b)** At the macroscopic level, misfolded α-synuclein propagates via structural connections, estimated from diffusion-weighted imaging. Simulated neuronal loss (atrophy) is compared against empirical atrophy, estimated from PD patients using deformation-based morphometry (DBM).

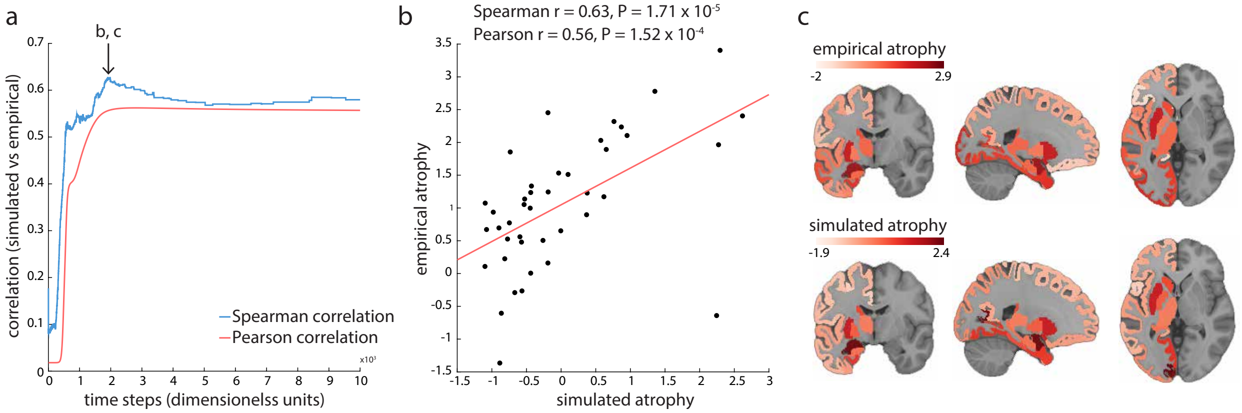

Figure 2: **Model fit** — **(a)** Correlations between simulated atrophy and empirical atrophy derived from PD patient DBM maps up to $t = 10^4$. Correlations are shown as a function of simulation time. After reaching the peak value ($r = 0.63, r = 1.71 \times 10^{-5}$), the model fit slightly drops and finally stabilizes. See Fig. S4 for correlations up to $t = 10^5$. **(b)** Model fit at the peak of Spearman's correlation taken from panel a. In scanning along t to find the peak value, early timeframes were discarded to avoid using a spurious correlation peak (see *Cutoff of the early-spreading timeframes* in *Appendices*). Using Pearson's correlation coefficients yielded comparable results ($r = 0.56, p = 1.52 \times 10^{-4}$). Values shown in the axes are normalized. The outlier at the bottom right is the nucleus accumbens (for a possible explanation see *Discussion*). **(c)** Simulated atrophy and empirical atrophy plotted on the ICBM152 standard MNI template. The slices were chosen at x=-22, y=-7, z=0 (MNI coordinates).

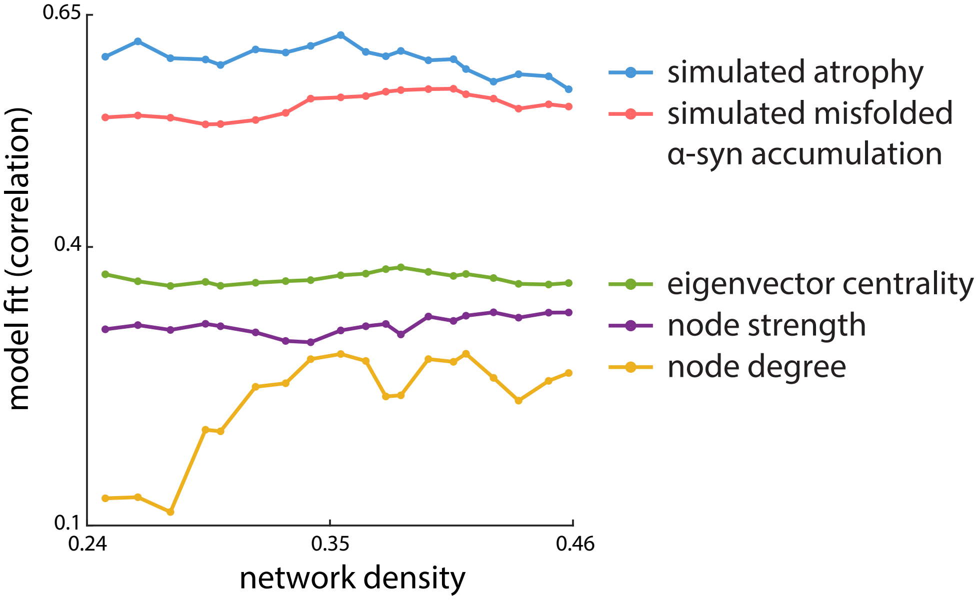

Figure 3: **The full dynamic model outperforms static network measures across multiple network densities** — The full spread model has more predictive power than static topological metrics, including node degree (yellow), node strength (purple) and eigenvector centrality (green). Moreover, simulated atrophy (blue) from the full agent-based model yielded higher correlation with empirical atrophy than the modeled density of misfolded α-synuclein (red, peak correlation along t at each density), suggesting that neuronal death resulting from misfolded α-synuclein accumulation plus deafferentation is a better measure of atrophy in PD than the mere accumulation of misfolded α-synuclein. Model fit was assessed using Spearman's correlation coefficient. The overall pattern of results was consistent across multiple network densities. Using Pearson's correlation coefficient yielded similar results (Fig. S2).

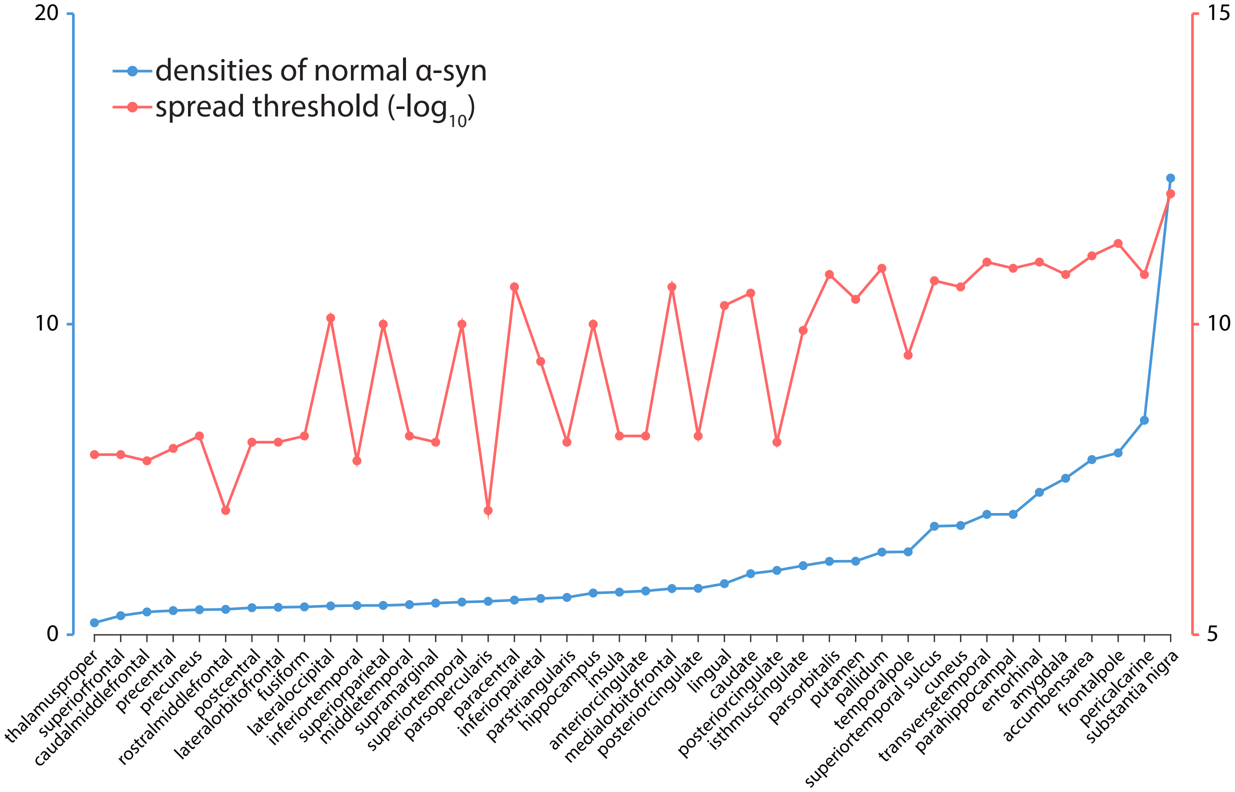

Figure 4: **Identifying the disease epicenter** — Densities of normal α-synuclein (blue) at equilibrium (represented by the stable point) and spread threshold (red). Spread threshold was inverted by $-\log_{10}$, so higher values indicate lower thresholds. Spread thresholds reflect the susceptibility of a region to trigger an epidemic. Basal ganglia regions are rich in endogenous α-synuclein (caudate ranks among the top 42.9% regions; putamen, 31.0%; pallidum, 28.6%) and have relatively low spread threshold (caudate ranks among the lowest 35.7%; putamen, 38.1%; pallidum, 16.7%). Substantia nigra has the highest normal α-synuclein level and lowest spread threshold, making it the most probable epicenter of disease outbreak.

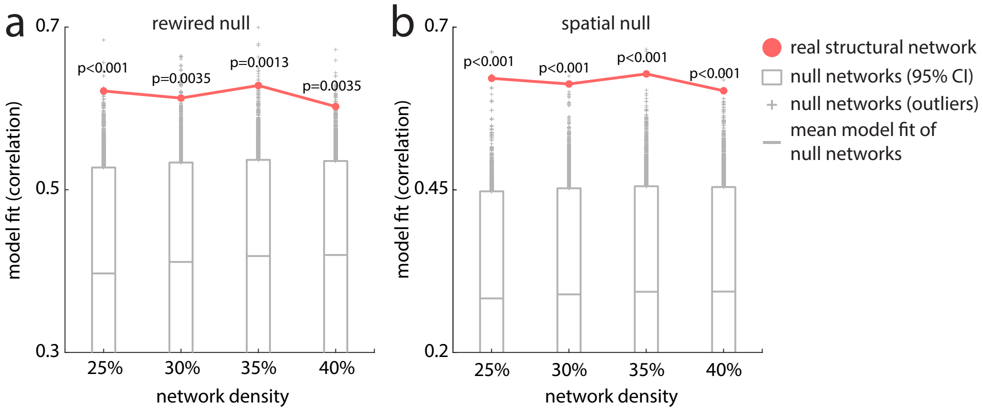

Figure 5: **Effects of network topology and geometry** — Systematic disruption of **(a)** connectome topology (rewired null) or **(b)** spatial embedding (spatial null) significantly degrades model fit as measured by Spearman's correlation. Red = real structural network (empirical network); grey = null networks. Rewired null: $p_{25\%} < 0.001$, $p_{30\%} = 0.0035$, $p_{35\%} = 0.0013$, $p_{40\%} = 0.0035$; spatial null: $p_{25\%} < 0.001$, $p_{30\%} < 0.001$, $p_{35\%} < 0.001$, $p_{40\%} < 0.001$).

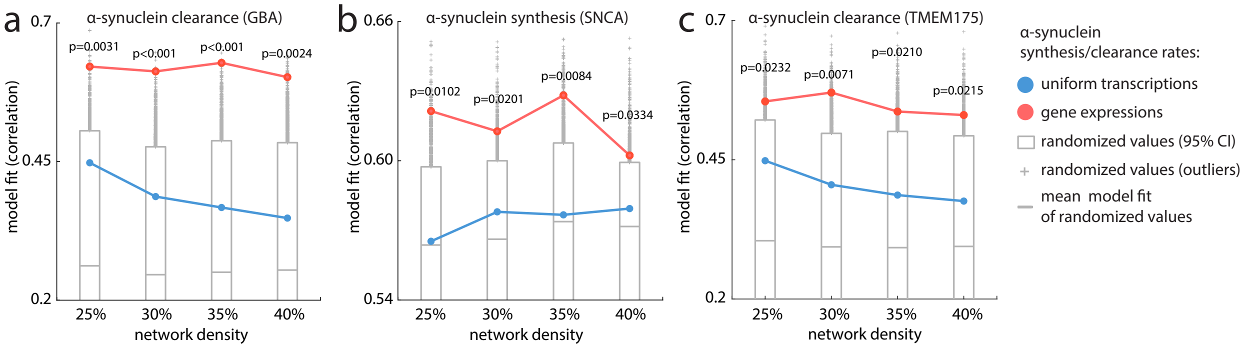

Figure 6: **Assessing the contribution of gene expressions** — To assess the influence of gene expression on atrophy, model fit using real expression values (red) is compared to null models in which node-wise expression profiles of *GBA* and *SNCA* (reflecting, respectively, α-synuclein clearance and synthesis) were shuffled. Both manipulations significantly reduce model fit regardless of network density (*GBA*: $p_{25\%} = 0.0031$, $p_{30\%} < 0.001$, $p_{35\%} < 0.001$, $p_{40\%} = 0.0024$; *SNCA*: $p_{25\%} = 0.0102$, $p_{30\%} = 0.0201$, $p_{35\%} = 0.0084$, $p_{40\%} = 0.0334$). Notably, uniform transcription profiles, in which all nodes have identical expression values (blue) yield above-chance model fit, but perform poorly compared to the model with real expression values (*GBA* uniform correlations: $r_{25\%} = 0.4479$, $r_{30\%} = 0.3869$, $r_{35\%} = 0.3672$, $r_{40\%} = 0.3481$; *SNCA* uniform correlations: $r_{25\%} = 0.5653$, $r_{30\%} = 0.5780$, $r_{35\%} = 0.5767$, $r_{40\%} = 0.5794$). Similar operations were conducted on *TMEM175*. Although the model with *TMEM175* coding for clearance rate yielded relatively lower model fit ($r_{25\%} = 0.4483$, $r_{30\%} = 0.4051$, $r_{35\%} = 0.3866$, $r_{40\%} = 0.3758$), null *TMEM175* expressions significantly degraded model fit as well ($p_{25\%} = 0.0232$, $p_{30\%} = 0.0071$, $p_{35\%} = 0.0210$, $p_{40\%} = 0.0215$).

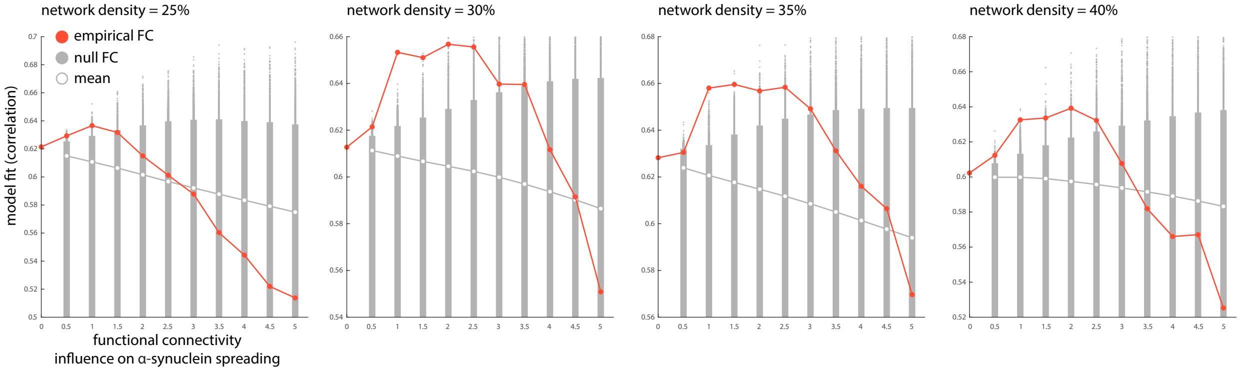

Figure 7: **Incorporating functional connectivity improves model fit** — Resting-state fMRI functional connectivity was incorporated in the model by tuning the probability of α-synuclein propagation along structural connections. As the influence of functional connectivity (k) is increased, α-synuclein spreading is biased towards structural connections that exhibit high functional connectivity. Model fit is shown for a range of structural connection densities. A balanced effect of functional connectivity and structural connectivity improves model performance, while excessive influence of functional connectivity degrades model fit. The same beneficial effect is not observed when randomized, "null" functional connectivity patterns are used (grey bar). At smaller k, simulations based on real functional connectivity yield significantly higher model fit than the null settings as indicated by the 95% confidence interval (gray bar), while at larger k, real functional connectivity ceases to have advantage over null functional connectivity matrices in facilitating model fit and can even become significantly more harmful than the nulls.

12 Appendices

12.1 Analysis of the fixed points

Although there is no analytical solution for α-synuclein concentration, phase plane analysis is helpful in finding fixed points of the system. Considering that the rates of incoming and outgoing agents in the edges are equal when the system is at the stable point, and that no clearance, synthesis, or misfolding occurs in the edges, the effects of the propagation module are negligible in the analysis of the system's fixed points. Therefore, we sought to use the "overall" or "total" synthesis rate (α), normal agent clearance rate (β_1), misfolded agent clearance rate (β_2), and transmission rate (γ) to approximately analyze the entire system using a series of differential equations (equations [15-17], see below). Likewise, we use N, M to represent the total population of normal and misfolded agents in the entire brain. β_1, β_2, γ depend on the actual N_i, M_i and thus are not static values (to see this, for example, the total cleared normal agents per unit time is $\sum_i \beta_i N_i$, so the "overall" clearance rate $\beta_1 = \sum_i \beta_i N_i / N$ depends on real-time N_i; it is not the "real" clearance rate but an approximation of the total rate of clearance); the total synthesis rate $\alpha = \sum_i \alpha_i S_i$, where α_i is the empirical synthesis rate in region i in the formal model and S_i is region size. Note that the actual spreading dynamics cannot be fully described by the following differential equations [15-17]. However, they are helpful in analyzing the possible states of disease propagation.

- **Growth of Normal α-synuclein.** The system of normal α-synuclein growth approximates like:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \alpha - \beta_1 N \tag{15}$$

  with $\alpha \gg 1, 0 < \beta_1 < 1$, obviously $N = \alpha / \beta_1$ is a stable point (which is the baseline density of edogenous α-synuclein, Fig. 4, blue line).

- **Spread of misfolded α-syn.** The system with misfolded agents injected behaves like:

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \alpha - \beta_1 N - (1-\beta_1)[1-(1-\gamma)^M]N \tag{16}$$

$$\frac{\mathrm{d}M}{\mathrm{d}t} = (1-\beta_1)[1-(1-\gamma)^M]N - \beta_2 M \tag{17}$$

The nullclines of N, M are

$$M = \frac{\ln(\frac{N-\alpha}{N(1-\beta_1)})}{\ln(1-\gamma)} \tag{18}$$

$$N = \frac{\beta_2 M}{(1-\beta_1)(1-(1-\gamma)^M)} \tag{19}$$

Note [19] is not defined at $M = 0$. To study the number of fixed points and their positions, we need to determine the number of intersections of the two nullclines [18] and [19], and where they intersect. Adding [16] and [17], it is easy to see $(N, M) = (\alpha/\beta_1, 0)$ is one fixed point. It is easy to find that M decreases monotonously with N in [18] and passes $(\alpha/\beta_1, 0)$. Therefore, the monotony and position of line [19] relative to line [18] becomes crucial. To find the monotony of [19], we take its first order derivative

$$N' = \frac{\beta_2(1-\beta_1)[(1-(1-\gamma)^M) + M(1-\gamma)^M \ln(1-\gamma)]}{[(1-\beta_1)(1-(1-\gamma)^M)]^2} \tag{20}$$

When $M = 0$, the first order derivative is 0. We then take the second derivative

$$N'' = C(\ln(1-\gamma))^2(1-\gamma)^M M \tag{21}$$

where C is a positive constant. When $M > 0$, [21] is positive thus the first order derivative [19] increases monotonously with M, and when $M < 0$, [19] decreases monotonously with M. Therefore, [20] is positive hence in [19], N increases monotonously with M, namely, there can be up to one intersection (denoted by (N^*, M^*) in the following) of the two nullclines apart from $(N, M) = (\alpha/\beta_1, 0)$. It determines that the system can have up to two fixed points, one is $(\alpha/\beta_1, 0)$, and the other (N^*, M^*), which has no closed-form expression.

It is also important to determine the position of (N^*, M^*), because $M* < 0$ is not realistic in the actual disease spread. Under this condition, an outbreak can never take place. Taking the limit of [19] $\lim_{M\to 0} N = \beta_2/(\beta_1-1)\ln(1-\gamma)$, we can see the intercept of [19] on the N axis is $\beta_2/(\beta_1-1)\ln(1-\gamma)$. When (i) $\beta_2 = (\beta_1-1)\ln(1-\gamma)\alpha/\beta_1$, $N^* = \alpha/\beta_1$, i.e., (N^*, M^*) "merges" with $(\alpha/\beta_1, 0)$; (ii) when $\beta_2 > (\beta_1-1)\ln(1-\gamma)\alpha/\beta_1 = 0$, $N^* > \alpha/\beta_1, M^* < 0$ which is not realistic; (iii) when $\beta_2 < (\beta_1-1)\ln(1-\gamma)\alpha/\beta_1 = 0$, $N^* < \alpha/\beta_1, M^* > 0$. Our results (with full spread in the end) are based on (iii).

To study why different choices of seed region and injected α-synuclein may lead to different states (extinction or outbreak), we also investigated under what conditions the fixed points are stable. This can be studied by taking the Jacobian matrix of the system linearized around the fixed points:

$$J = \begin{pmatrix} -\beta_1 - (1-\beta_1)[1-(1-\gamma)^M] & (1-\beta_1)(1-\gamma)^M \ln(1-\gamma)N \\ (1-\beta_1)[1-(1-\gamma)^M] & -\beta_2 - (1-\beta_1)(1-\gamma)^M \ln(1-\gamma)N \end{pmatrix} \tag{22}$$

At $(\alpha/\beta_1, 0)$, $J|_{(\alpha/\beta_1,0)} = \begin{pmatrix} -\beta_1 & (1-\beta_1)\ln(1-\gamma)\frac{\alpha}{\beta_1} \\ 0 & -\beta_2 - (1-\beta_1)\ln(1-\gamma)\frac{\alpha}{\beta_1} \end{pmatrix}$ has eigenvalues $-\beta_1$ and $-\beta_2 - (1-\beta_1)\ln(1-\gamma)\alpha/\beta_1$. When $\beta_2 > (\beta_1-1)\ln(1-\gamma)\alpha/\beta_1$, both eigenvalues are negative hence $(\alpha/\beta_1, 0)$ is stable (disease extinction). The injection of misfolded agents introduces a small perturbation to the system at $(\alpha/\beta_1, 0)$. As the choice of seed region and injection amount affect the approximations of β_1, β_2, γ in equation [16-17] that are used to analyze the system, the disease will either die out or fully spread. It is more difficult to initiate the disease spread in seed regions with relatively large

β_1, β_2 and small γ (i.e., more resistant to disease spread), as it is more likely to satisfy the condition $\beta_2 > (\beta_1 - 1)\ln(1-\gamma)\alpha/\beta_1$ initially. As mentioned before, β_1, β_2, γ are not static and depend on the real-time N_i, M_is; at the other fixed point (N^*, M^*), the parameter set is not the same as the one near $(\alpha/\beta_1, 0)$. Therefore it is also possible, in theory, that certain choices of parameters may lead to an outbreak followed by gradual extinction.

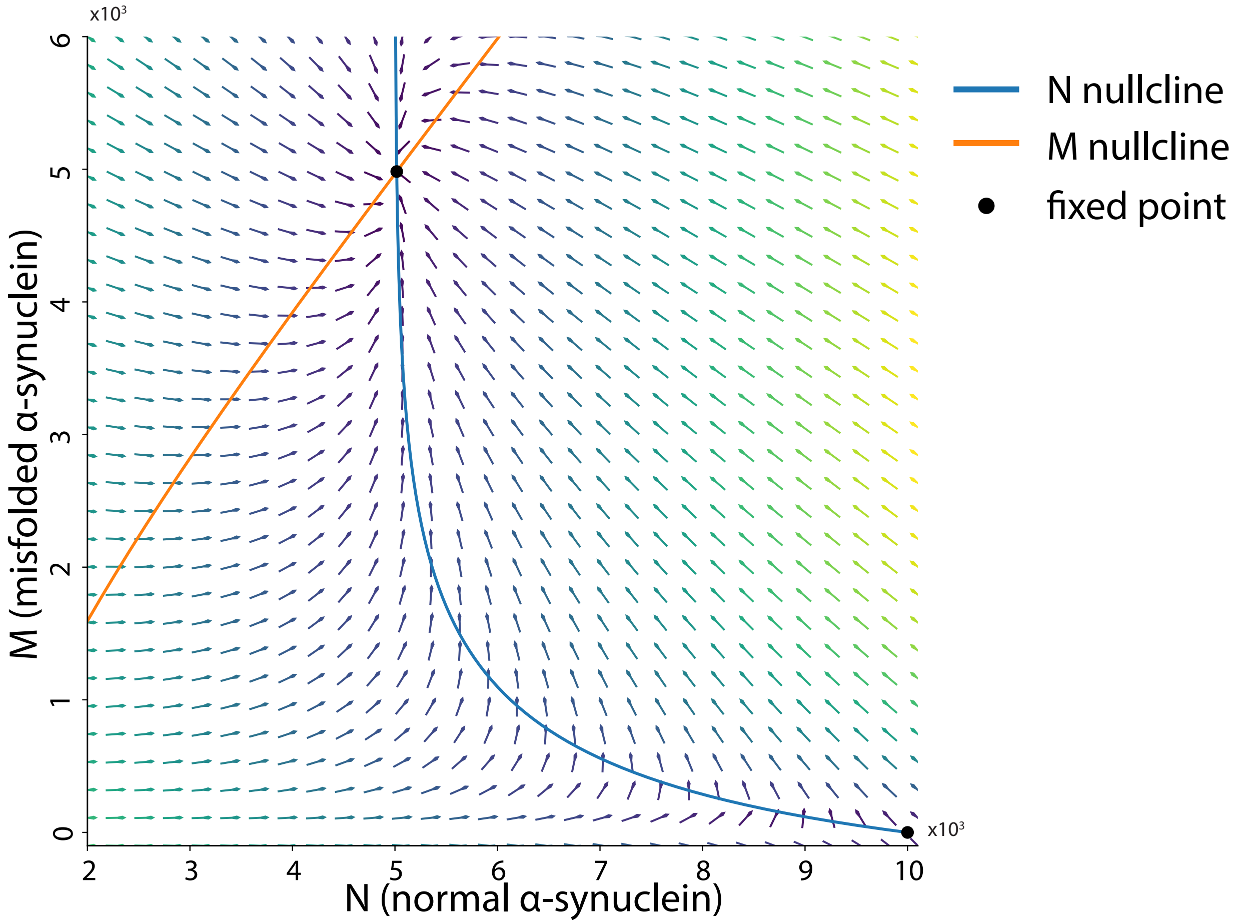

Figure S1: **An illustration of the phase plane at $\alpha = 5000, \beta_1 = 0.5, \beta_2 = 0.5, \gamma = 0.001$.** — M decreases with N (N nullcline, blue, equation [18]) and N increases with M (M nullcline, orange, equation [19]), therefore apart from $(N = 10000, M = 0)$ there is only one another intersection $(N = 5017.15, M = 4982.85)$ of the two lines, indicating that the system has two fixed points only. The vector field (arrows) denotes the direction of the gradient at each position (i.e., the system at that point will move along the direction of the corresponding arrow).

12.2 List of Notations

notation	**name**	**expression or value**	**explanation**
Δt	time step	$\Delta t = 0.01$	time increment in the simulations
i	region index	N/A	used to index regions
(i, j)	edge label	N/A	used to index edges
N_i	normal population in region i	N/A	total number of normal agents in region i
M_i	misfolded population in region i	N/A	total number of misfolded agents in region i
$N_{(i,j)}$	normal population in edge (i, j)	N/A	total number of normal agents in edge (i, j)
$M_{(i,j)}$	misfolded population in edge (i, j)	N/A	total number of misfolded agents in edge (i, j)
α_i	synthesis rate in region i	$\alpha_i = \Phi_{0,1}(SNCA\text{expression}_i)$	the probability that a new normal agent gets synthesized in each voxel of region i per unit time
β_i	clearance rate in region i	$\alpha_i = \Phi_{0,1}(GBA\text{expression}_i)$	the probability that an existing agent (either normal or misfolded) in region i gets cleared per unit time
S_i	region size	N/A	voxel counts of region i
γ_i^0	baseline transmission rate	$\gamma_i^0 = 1/S_i$	the probability for a single misfolded agent to transmit the disease to other agents per unit time
γ_i	transmission probability	$1 - \mathrm{e}^{M_i \ln(1-\gamma_i^0)}$	the probability that normal agents get infected (by at least one of the misfolded agents) per unit time
w_{ij}	connection strength of edge (i, j)	normalized fiber tracts density between region i and j	determining the probability of choosing edge (i, j) when exiting region i per unit time
l_{ij}	edge length of edge (i, j)	mean length of fiber tracts between region i and j	determining the probability of exiting edge (i, j) per unit time
ρ_i	the probability of remaining in region i	$\rho_i = 0.5$ for all i	Agents in region i have equal probability of remaining in region i or exiting region i per unit time
fc_{ij}	functional connectivity of edge (i, j)	N/A	biasing agents toward regions showing greater co-activation pattern
k	weight of functional connecitivity	N/A	controling the influence of functional connectivity in driving disease spread
k_1	weight of atrophy accrual due to native misfolded agents' accumulation	$k_1 + k_2 = 1$	controling the contribution of native misfolded agents' accumulation to total atrophy growth
k_2	weight of atrophy accrual due to deafferentation	$k_1 + k_2 = 1$	controling the contribution of deafferentation to total atrophy growth.
$r_i(t)$	the ratio of misfolded agents in region i	$r_i(t) = M_i(t)/(N_i(t) + M_i(t))$	measuring the burden of misfolded agents in region i at t. $1 - \mathrm{e}^{-r_i(t)\Delta t}$ is the increment of neuronal death at t in region i

Table 1: **Notation List** — Note that only k, ρ_i, k_1, k_2 are free parameters: k was scanned from 0 to 5 to study the effect of functional connectivity on disease spread (FIG. 7); $\rho_i = 0.5$ for all the regions so that agents have equal chance of staying in the same region or moving out; $k_1 = k_2 = 0.5$ so that the two factors ((i) native misfolded α-synuclein accumulation; (ii) deafferentation from connected regions) contributed equally to the total atrophy growth. We also note that model fit is robust across multiple choices of ρ_i, k_1, k_2 (FIG. S7).

12.3 Spearman's correlation vs Pearson's Correlation

. We adopted Spearman's correlation to assess model fit instead of Pearson's correlation. Although assessing model fit using Spearman's correlation takes into account only relative magnitudes of simulated neuronal loss and inevitably discards the information of data points' values, thus can be sensitive to small changes that alter the rank order of regional neuronal loss, it is better capable of capturing the similarity of the rank orders between two variables. It is preferred in our present study as it measures the resemblance between the simulated atrophy and empirical atrophy as to which region(s) display more atrophy compared to other regions. Moreover, the simulated neuronal loss does not exhibit a normal distribution pattern, making the Pearson's correlation less suitable in out study. However, to ensure the robustness of model fit, we also derived Pearson's correlations across the same set of network densities, which all yield comparable results (FIG. S2). Also note that, considering that the low spatial resolution of structural MRI scans of PD patients may cause inaccuracy in assessing the atrophy in substantia nigra, it was excluded in the correlations.

12.4 Cutoff of the early-spreading timeframes

The sensitivity of Spearman's correlation to rank orders may be problematic in the early spreading period, as the measure takes in ranking information only so that even small increments that alter ranking order of the original neuronal loss may cause substantial changes to model fit. Therefore, to avoid picking up spurious peak correlation value (model fit) in the early timeframes after seeding, we discarded the timeframes where change of misfolded α-synuclein densities exceeds 1% within $\Delta t = 0.01$ in at least one region, resulting in a cutoff point at around 1000 time steps depending on the network density.

We also adopted a less rigorous cutoff point, removing only the first 100 timeframes. The difference in results is negligible (FIG. S3).

12.5 Supplementary figures

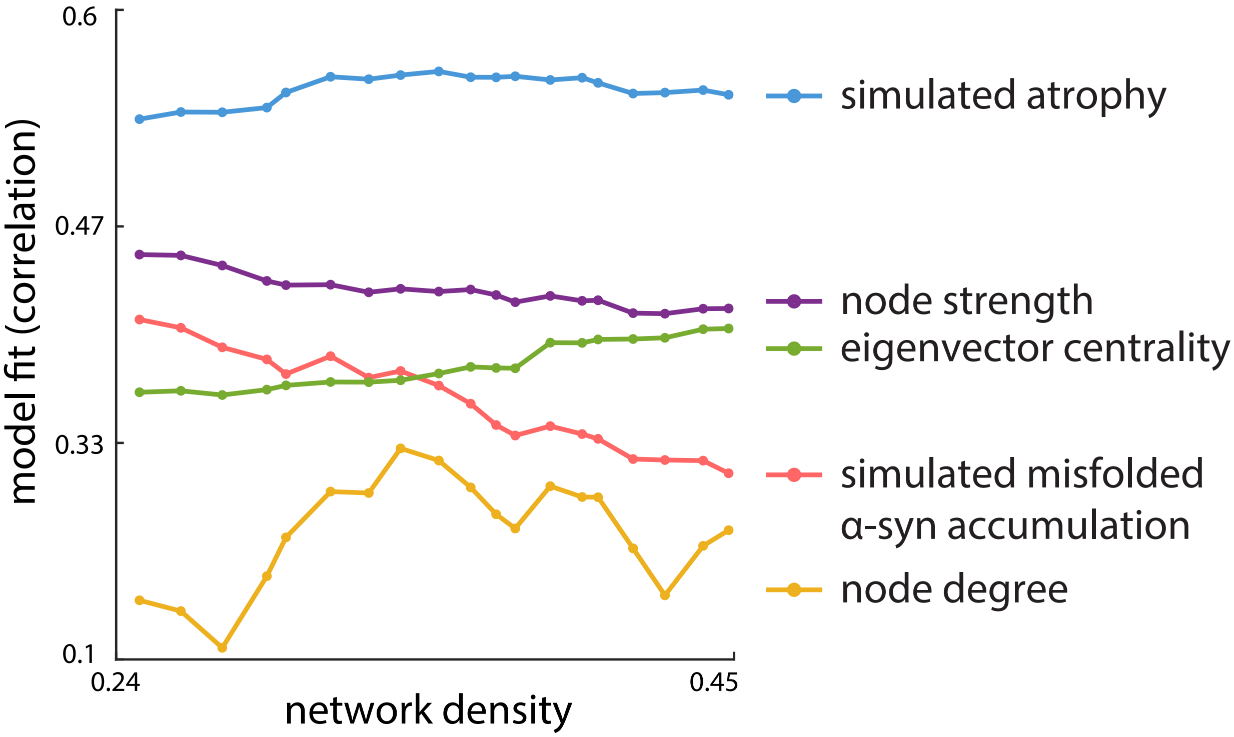

Figure S2: **Model fit based on Pearson's correlation coefficient yielded comparable results across network density from 25% to 45%** — The model integrated with gene expression levels has more predictive power than the density of misfolded α-synuclein (red) and the static network metrics, including node degree (yellow), node strength (green), or eigenvector (purple) centrality.

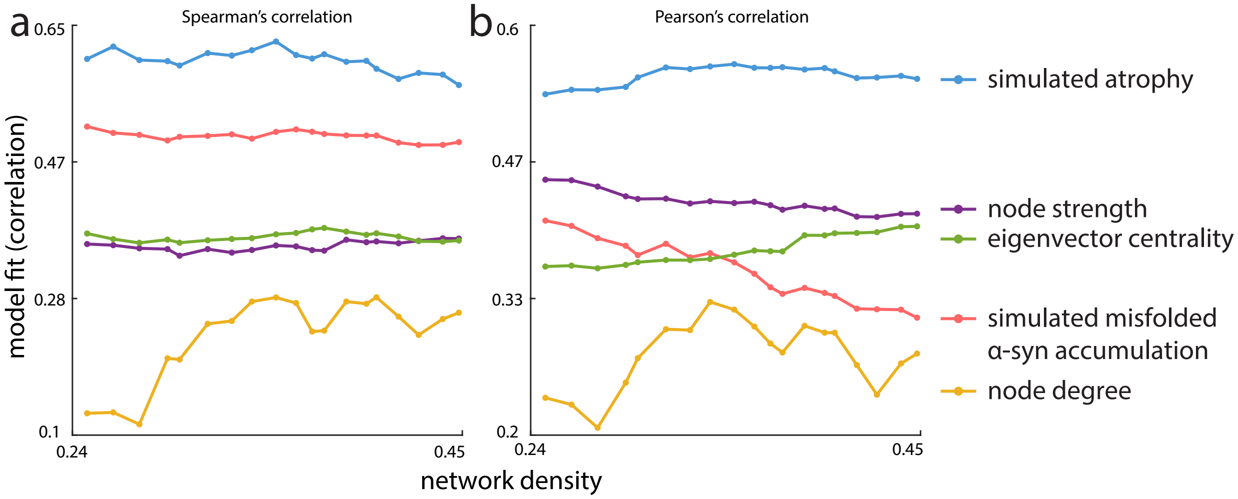

Figure S3: **The model fit across network density 25% to 45% after removing the first 100 timeframes in finding the peak correlation values.** — **(a)** Spearman's correlation. **(b)** Pearson's correlation.

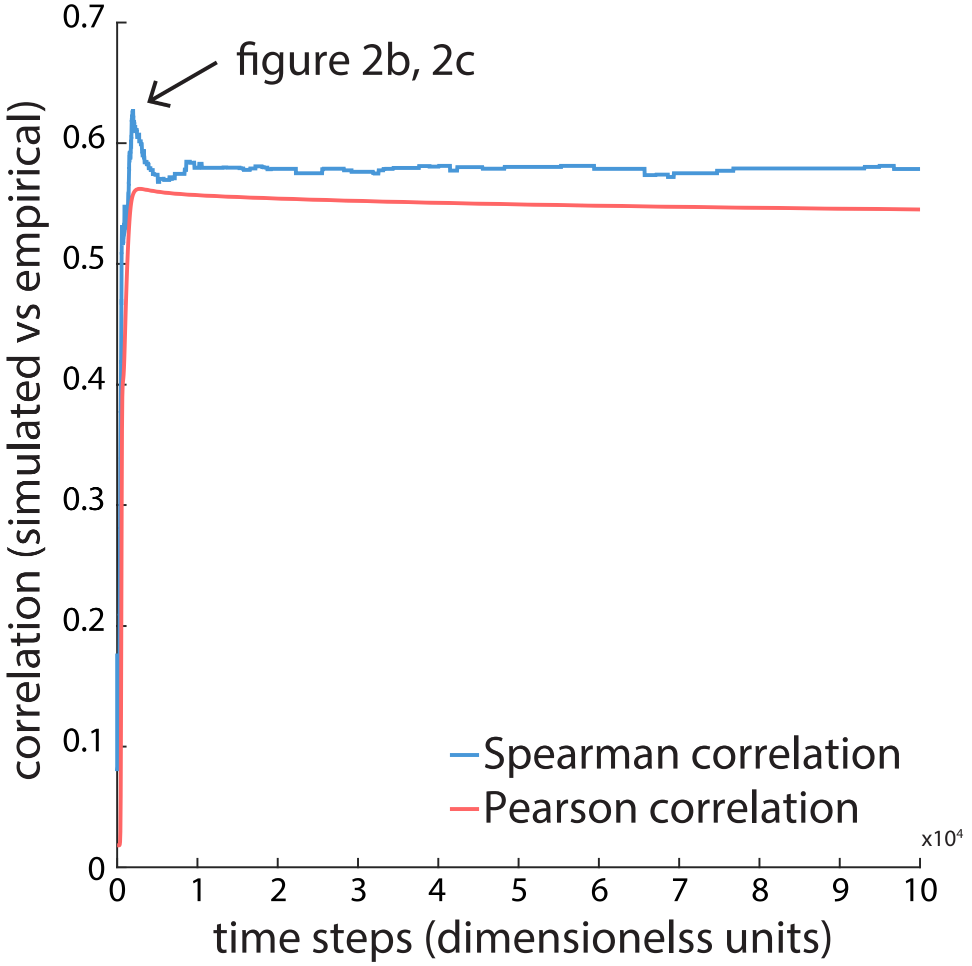

Figure S4: **Model fit up to $t = 10^5$** — Correlations between simulated atrophy and empirical atrophy derived from PD patient DBM maps. Correlations are shown as a function of simulation time. At large t, the model fit stabilizes as the system approaches the stable point.

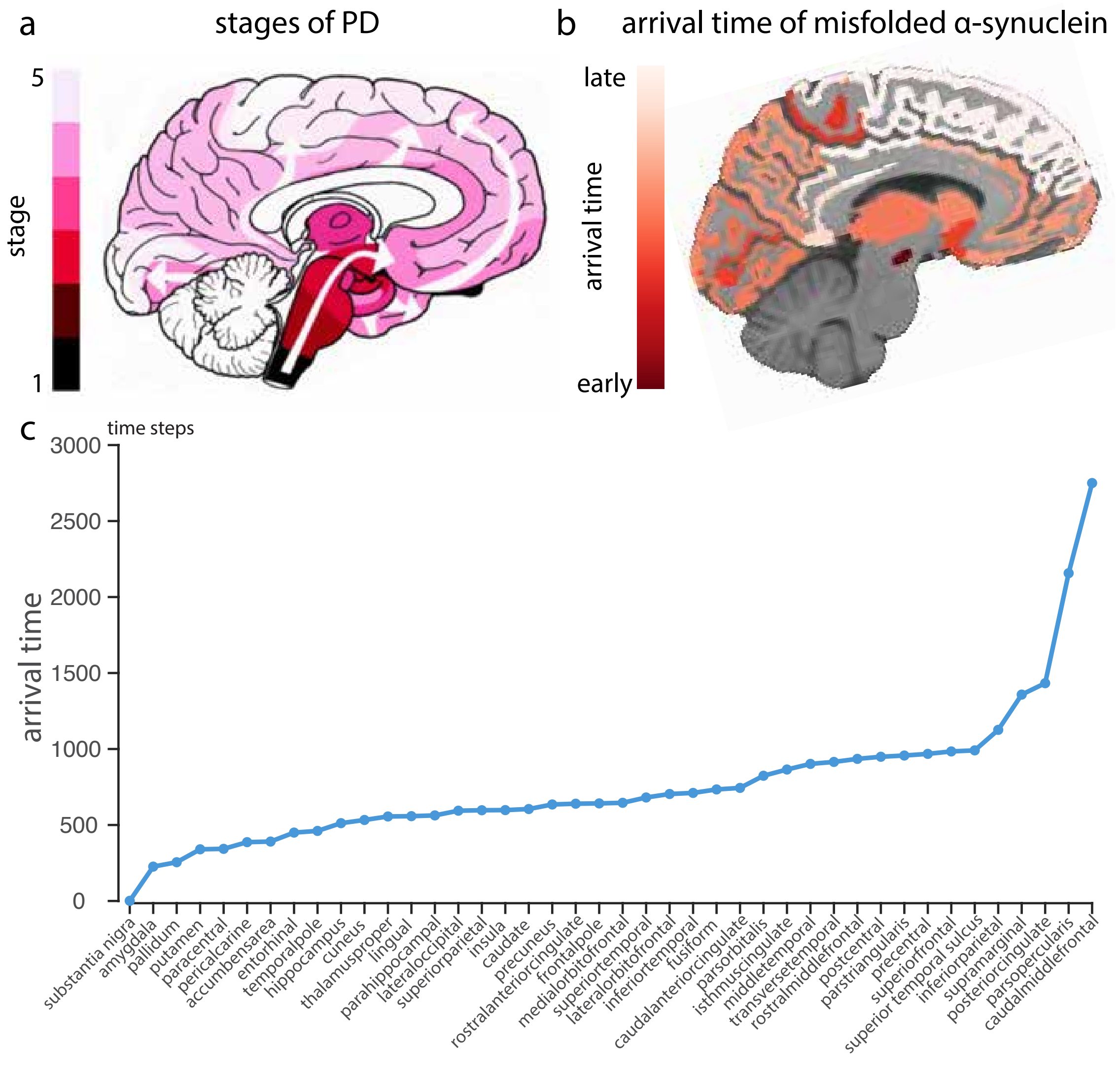

Figure S5: **Comparison between stages of PD and arrival time of misfolded α-synuclein in the model.** (a) The stages of PD progression proposed by Braak. (b) Regional arrival time of misfolded α-synuclein is defined as the time steps required for misfolded α-synuclein amount to exceed 1 (after seeding at the substantia nigra with one misfolded agent). This roughly follows the Braak staging hypothesis. (c) Line chart of regional arrival time of misfolded α-synuclein.

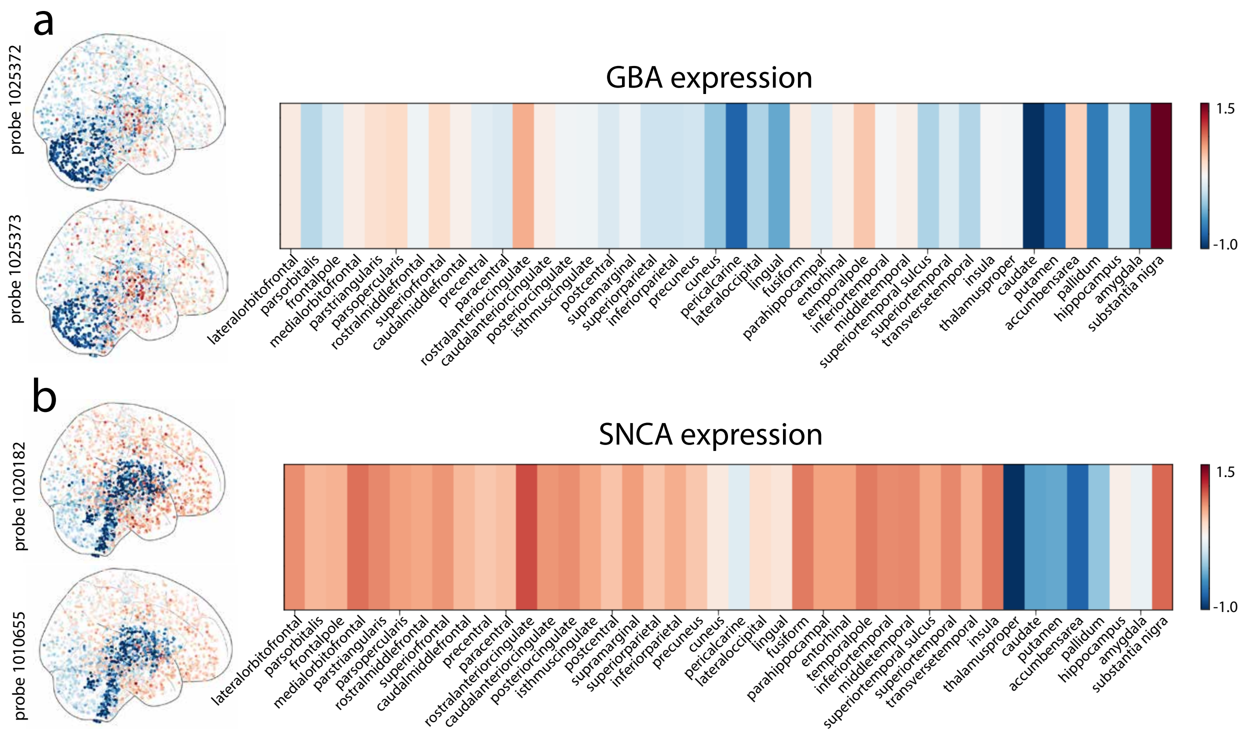

Figure S6: (a) **Regional GBA expression.** There are three probes for *GBA* (probe id: 1025372, 1025373, 1025374). Probe 1025372, 1025373 were included to generate the group transcription profile. *GBA* probe 1025374 was excluded in the analysis as it deviates too much from probe 1025372 (Pearson correlation=0.30) and probe 1025373 (Pearson correlation=0.24) while the correlation between the latter two probes is 0.79. (b) **Regional SNCA expression.** There are two probes for *SNCA* (probe id: 1020182, 1010655). Boths probes were included to generate the group transcription profile. Compared to *GBA* expression, *SNCA* is more homogeneous especially in cortical regions.

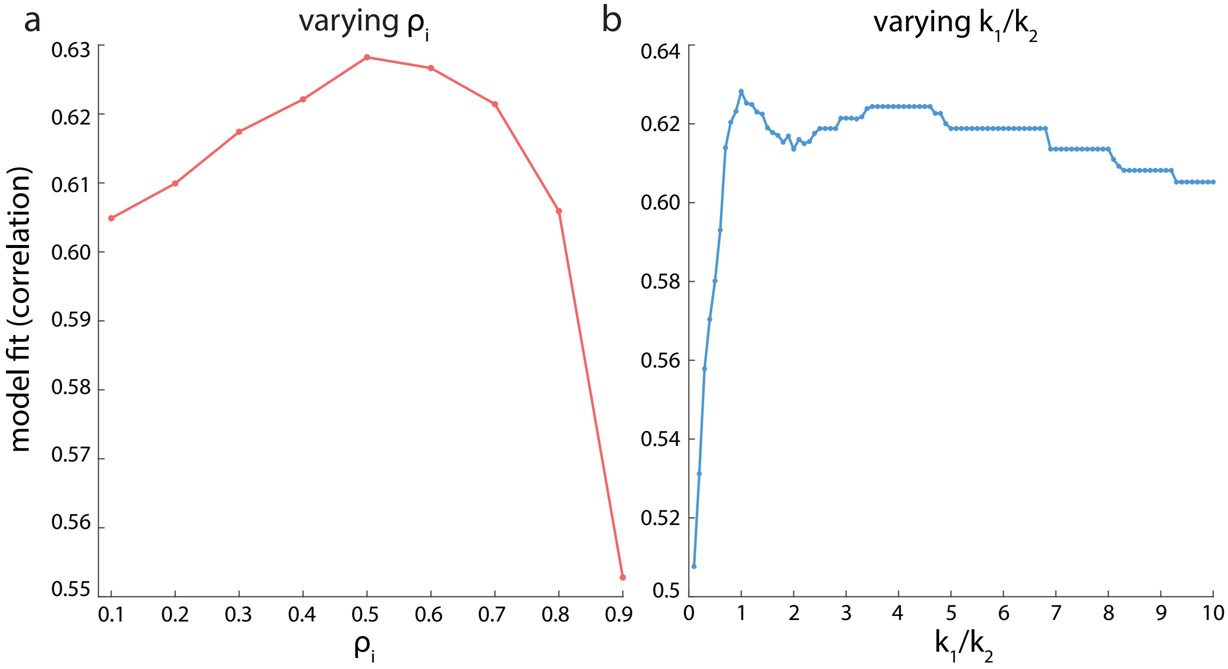

Figure S7: **Testing free parameters** ρ_i, k_1, k_2 — Model fit (Spearman's correlation) is robust to variations in ρ_i, k_1, k_2 (results shown at network density %35). **(a)** ρ_i controls the probability of remaining in region i while (1-ρ_i) is the probability of exiting region i per unit time. The main results are based on $\rho_i = 0.5$. However, the model fit is consistently above 0.55 across ρ_i ranging from 0.1 to 0.9. **(b)** For the atrophy in region i, k_1 controls the contribution of α-synuclein accumulation inside region i, while k_2 controls the contribution of deafferentation induced by atrophy in connected regions. $k_1 + k_2 = 1$. The model fit is consistently over 0.5 across k_1/k_2 ranging from 0.1 to 10. These results suggest that the predicative power of the model is robust to variations in free parameters ρ_i or k_1/k_2.