bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Local vulnerability and global connectivity jointly shape neurodegenerative disease propagation

Ying-Qiu Zheng[1], Yu Zhang[1,2], Yvonne Yau[1], Yahar Zeighami[1], Kevin Larcher[1], Bratislav Misic[1◑*], Alain Dagher[1◑*]

1 McConnell Brain Imaging Centre, Montréal Neurological Institute, McGill University, Montréal, Quebec, Canada
2 Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, Canada

◑These authors contributed equally to this work.
* bratislav.misic@mcgill.ca, alain.dagher@mcgill.ca

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 and prion-like propagation of endogenous α-synuclein 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. The dynamic model replicates the spatial and temporal patterning of empirical atrophy in PD and implicates the substantia nigra as the disease epicenter. We reveal a significant role for both connectome topology and geometry in shaping the distribution of atrophy. The model also demonstrates that *SNCA* and *GBA* transcription influence α-synuclein concentration and local regional vulnerability. Functional co-activation further amplifies the course set by connectome architecture and gene expression. Altogether, these results support the theory that the progression of PD is a multifactorial process that depends on both cell-to-cell spreading of misfolded proteins and regional vulnerability.

Introduction

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 [1–6]. 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

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 [7], and direct evidence in humans remains mostly circumstantial [8].

The prion hypothesis suggests that propagation dynamics in neurodegenerative diseases may be modeled using methods derived from infectious disease epidemiology. Just as infectious diseases spread via social contact networks, misfolded proteins propagate via the brain's connectome. There are different approaches for modeling epidemic spread over a network. 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 [9, 10].

Such models are easily solved mathematically but have limited explanatory power. Another approach is the agent-based model [11], 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. 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).

Here we propose a Susceptible-Infected-Removed (S-I-R) agent-based model on a brain network to explore 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. Although convincing evidence from animal [12–18] and neuroimaging studies [19, 20] supports the propagation of misfolded and neurotoxic α-synuclein, other mechanisms may also drive PD pathology, including cell-autonomous factors, dependent on gene expression, that modulate regional neuronal vulnerability [7]. 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. We also investigate whether selective vulnerability may influence the spatial patterning of the disease.

Fig 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).

Results

Model construction

- *Structural connectivity.* Diffusion-weighted MRI data from N=1027 healthy participants was used to construct the anatomical network for α-synuclein

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

propagation (source: Human Connectome Project, 2017 S1200 release; [21]). Adjacency matrices were reconstructed using deterministic streamline tractography [22]. 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% [23–25].

- *Functional connectivity.* Resting-state functional MRI (fMRI) data from N=496 healthy participants (source: Human Connectome Project, 2015 S500 release; [21]) 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 Allen Human Brain Atlas (AHBA) dataset. These gene expression profiles determine the local synthesis and degradation of α-synuclein (see *Methods*).

- *Atrophy.* An atrophy map was derived from T1-weighted MRI scans of 237 PD patients and 118 age-matched healthy controls (source: Parkinson progression marker initiative (PPMI); [26]). 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 [19].

- *Brain parcellation* The brain MRI template was parcellated according to an anatomical segmentation-based atlas, featuring 68 bilateral cortical and 15 subcortical regions [27–29]. 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, 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 the 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

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

distribution function. The system has only one stable point which can be found numerically (see S1 Text and S1 Fig), 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. 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 (see S1 Text and S1 Fig). 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 S1 Table for the full list of parameters and their explanations.

Simulated neuronal loss replicates the spatial pattern of atrophy

We first investigated whether misfolded α-synuclein spreading on the healthy connectome could 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 or synaptic loss directly; (2) atrophy due to deafferentation secondary to cell death or synaptic loss 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 2A) after which it drops slightly and stabilizes (see S1 Text). 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 2B shows the linear relationship between simulated and empirical atrophy across all nodes at peak fit, while Fig 2C 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 S2 Fig for model fit up to 10^5 time steps), a finding that mirrors recent discoveries in animal models where misfolded α-synuclein eventually ceases to increase in later stages [30]. We also note that misfolded α-synuclein arrival time at each brain region resembles the

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig 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 S2 Fig for correlations up to $t = 10^5$. (B) Model fit at the peak of Spearman's correlation taken from panel a. 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). The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/blob/master/results/Fig2.mat.

well-established Braak stages of PD [31, 32] (S3 Fig). For validation purposes, we estimated the DBM values using an alternative pipeline (*fsl_anat*) [33], re-obtained the t-statistics as the atrophy measure, and found the model fit based on the new measure yielded a comparable temporal pattern (S4 Fig).

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 atrophy estimate 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 tissue loss due to both α-synuclein and deafferentation yields a better model of regional atrophy accrual than the mere accumulation of misfolded α-synuclein. Note that as Spearman's correlation is relatively unstable when sample size is limited, it may peak at early-spreading time frames while at the time the simulated atrophy bears no real resemblance to the real atrophy pattern. At the same time, atrophy is a late stage in symptom progression of PD. We therefore discarded the early-spreading timeframes, defined as the timesteps at which change of misfolded α-synuclein density in any of the regions exceeds 1%.

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 [34]. 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

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig 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 loss of neuronal tissue 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 (S5 Fig). For the same analysis using two finer-grained anatomical parcellations see S6 Fig. The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/blob/master/results/Fig3.mat.

model provide explanatory power above and beyond network topology and gene expression.

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 [30, 35]; likewise, in our model, higher regional abundance of normal α-synuclein agents should indicate 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 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 observations of Lewy body inclusions and dopaminergic neuron loss in substantia nigra of PD patients as well as its role in most of the presenting symptoms of the disease [32, 36, 37]. 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 [20]. 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 propagate an outbreak. As explained in the previous section, the agent-based model has two fixed points representing disease extinction or major outbreak. Although, in our model, the choice of seed region and injection number of misfolded α-synuclein agents does not affect the final magnitude of misfolded α-synuclein accumulation, it can 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 in a brain region indicates its likelihood of acting as an epicenter. Therefore, we quantified the spread threshold for

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig 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. The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/blob/master/results/Fig4.mat.

each 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 will be affected by an infectious agent before it is removed) marks the transition between the regimes in which disease spreads or extinguishes [38]. 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 by scanning across different injected amounts of misfolded α-synuclein to find the point at which the disease no longer extinguishes. More specifically, starting with an injection amount at which the disease does not spread (here we chose 1e-13), we incremented the value by step sizes of 1e-13 until the point where the disease no longer extinguishes, and took this as the spread threshold. This procedure was repeated for every region, yielding 42 regional spread thresholds (Fig 4, red curve).

Substantia nigra has the lowest spread threshold (Fig 4, red curve), 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 [19], and is generally one of the earliest regions to display neuronal loss in clinically overt 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, which are likely affected earlier than the substantia nigra (see *Discussion*).

Connectome architecture shapes disease spread

We next asked whether model fit depends on the connectome's topology and/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) (Fig. 5).

Fig 5. Effects of network topology and geometry. (A) Systematic disruption of connectome topology (rewired null). (B) Systematic disruption of spatial embedding (spatial null). Both procedures significantly degrade 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: $p_{25\%} < 0.001$, $p_{30\%} < 0.001$, $p_{35\%} < 0.001$, $p_{40\%} < 0.001$). The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/blob/master/results/Fig5.mat.

Rewired null networks were generated by swapping pairs of edges while preserving the original degree sequence and density using the Maslov-Sneppen algorithm [39]

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

implemented in the Brain Connectivity Toolbox (https://sites.google.com/site/bctnet/) [40]. 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 [41, 42]. 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 5A: $p_{25\%} < 0.001$, $p_{30\%} = 0.0035$, $p_{35\%} = 0.0013$, $p_{40\%} = 0.0035$; spatial null, Fig 5B: $p_{25\%} < 0.001$, $p_{30\%} < 0.001$, $p_{35\%} < 0.001$, $p_{40\%} < 0.001$) and suggests that the high correspondence between simulated and empirical atrophy in PD is jointly driven by connectome topology and geometry.

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. (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 6A and 6B 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 curve) and permuted gene expression levels (Fig 6, grey bar).

Shuffling the transcription profile of either gene significantly degraded model fit (Fig 6A, *GBA*: $p_{25\%} = 0.0031$, $p_{30\%} < 0.001$, $p_{35\%} < 0.001$, $p_{40\%} = 0.0024$; Fig 6B, *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.

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

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fig 6. Assessing the contribution of *GBA* and *SNCA* gene expression. 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. (A) Model fit of randomized *GBA* expression (grey bar) is significantly worse than that of the real *GBA* expression (red line). $p_{25\%} = 0.0031$, $p_{30\%} < 0.001$, $p_{35\%} < 0.001$, $p_{40\%} = 0.0024$. (B) Model fit of randomized *SNCA* expression (grey) is significantly worse than that of the real *SNCA* expression (red). $p_{25\%} = 0.0102$, $p_{30\%} = 0.0201$, $p_{35\%} = 0.0084$, $p_{40\%} = 0.0334$). Notably, uniform transcription profiles, in which all nodes have identical expression values (blue line) 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: $r_{25\%} = 0.5653$, $r_{30\%} = 0.5780$, $r_{35\%} = 0.5767$, $r_{40\%} = 0.5794$). The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/blob/master/results/Fig6.mat.

underperformed compared to those using empirical transcription profiles (Fig. 6, red = empirical gene expression, blue = uniform gene expression); 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, here dependent on α-synuclein concentration. 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: $r_{25\%} = 0.5653$, $r_{30\%} = 0.5780$, $r_{35\%} = 0.5767$, $r_{40\%} = 0.5794$, blue curve 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.

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 [19, 20]. Secretion of α-synuclein by neurons has been shown to be activity dependent [43]. 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 (Fig 7A). 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.

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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). 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.

Fig 7. Incorporating functional connectivity improves model fit. (A) An illustration of incorporating the influence of functional connectivity. Region 1 is more densely connected with region 2 than with region 3 (i.e., structural connectivity $w_{12} > w_{13}$) but co-activates more with region 3 than with region 2 (i.e., functional connectivity $\mathrm{fc}_{12} < \mathrm{fc}_{13}$). If functional connectivity is not incorporated, the probability of spreading towards region 2 or 3 for agents in region 1 is proportional to the structural connectivity w_{12} or w_{13} (upper panel); after functional connectivity is incorporated, these probabilities are rescaled to be proportional to $\exp(k \times \mathrm{fc}_{12})w_{12}$ and $\exp(k \times \mathrm{fc}_{13})w_{13}$ respectively (lower panel) where k is a factor to control the importance of functional connectivity. (B) 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 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 (S7 Fig). The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/blob/master/results/Fig7.mat.

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. The results are shown in S7 Fig. 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 supports the conclusion that atrophy patterns observed in PD patients depend on both the structural and functional architecture of the brain.

Discussion

Modeling the spatio-temporal dynamics of neurodegeneration

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

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

incorporates pathogenic spread (dominated by the connectome) and selective vulnerability (modulated here by α-synuclein concentration derived from gene expression) under one 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.

The S-I-R model allowed 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 (S1 Text, S1 Fig). If the system is attracted to extinction, misfolded proteins will eventually be cleared. If the system is attracted to the second fixed point, this represents an outbreak. However, in our model, misfolded proteins do not accumulate boundlessly, but achieve a stable final concentration where they co-exist with normal proteins. These results are consistent with recent experimental evidence in rodents where injected misfolded α-synuclein grew but eventually ceased to propagate [30], 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 [44].

Agent-based models built on networks allow for the determination of the origin of a disease outbreak and arrival times at different locations [45]. Combining network structure with protein spreading dynamics allowed for the identification of the substantia nigra as the likeliest disease epicenter. Mirroring the *Reproduction number* R_0 [38], 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 to cause an outbreak. In our model, the substantia nigra has the lowest spread threshold, identifying it as a likely disease epicenter. This is not to say that the substantia nigra is the origin of the disease, or the first affected site: the concept of epicenter as used here is similar to "best propagator" [10], and identifies the region most likely to trigger an outbreak, rather than the first affected site. According to the Braak hypothesis, the dorsal motor nucleus of the vagus is the initial central nervous system target in PD [31, 32]; however, we could not include structures in the pons and medulla due to difficulty in imaging either atrophy or white matter tracts in the brainstem. Nonetheless, our model is consistent with the substantia nigra acting as a propagator of disease from brainstem to supratentorial areas [19]. We suggest that this may result from its high concentrations of α-synuclein and widespread connections. We also used the agent-based model to estimate α-synuclein arrival time at each brain region after seeding the substantia nigra (S3 Fig).

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 [9, 10] and to neurodegeneration more generally [46]; 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. This simplified model can easily be tailored to accommodate a full agent-based setting by

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

Interplay of local vulnerability and network propagation

The small-world properties of brain networks that favour information flow may also potentiate disease spread [47]. These properties include short path lengths [48] and community structure [49, 50], the second of which may potentiate global disease spread by enhancing local, intra-community infection [51]. The presence of high degree nodes (hubs) that are highly interconnected also favours disease propagation [52]. Hubs 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 [34], including PD [19]. Here we show that disruptions of brain network architecture reduce model fit, providing evidence that the emergent dynamics of synucleinopathy depend on network topology and geometry.

However, 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 of α-synuclein both impose constraints on the spreading process. 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. More specifically, the agent-based model allowed us to test two competing theories of PD pathogenesis: prion-like protein propagation versus regional vulnerability [1, 7]. Here we chose to model regional vulnerability by incorporating estimated local α-synuclein concentration, known to facilitate seeding [53] and increase neuronal vulnerability in animal models [35]. 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 [54]. Mutations in *GBA* are the most common genetic risk factor for PD [55, 56]; mutations and multiplications of *SNCA* have been implicated in driving the severity of the pathology [57–59]. It is worth noting that simple spatial correlation measures alone failed to link *GBA* or *SNCA* regional expression to empirical atrophy; 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 for the PD model, or to adapt it to other neurodegenerative diseases.

It is also known that α-synuclein is secreted in an activity dependent manner [43]. 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

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

propagation, explaining why atrophy patterns in neurodegenerative diseases tend to resemble intrinsic functional networks [19, 60].

Methodological considerations

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 apart from the effects of α-synuclein concentration were not accounted for. 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 [61]. Moreover, tissue loss was homogeneously modelled as a simple linear combination of local damage (from α-synuclein accumulation) and deafferentation, which may not reflect reality. Also, 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.

Note also that our model does not attempt to distinguish between neuronal, axonal, dendritic or other tissue loss as causes of atrophy. Our only hypothesis is that tissue damage from α-synuclein accumulation is reflected in MRI deformation. However, atrophy measured with DBM does not necessarily reflect death of neurons. Indeed, postmortem studies in PD demonstrate a dissociation between alpha-synuclein pathology and neuronal loss, which is prominent in some areas (e.g. substantia nigra), but virtually absent in others (cortex, amygdala) [72]. This is similar to normal senescence, where there is widespread tissue atrophy in cortex despite preservation of neuronal numbers [73]. Thus, it is possible that tissue atrophy in PD in some regions may reflect loss of dendritic arbors and spines without loss of neurons.

Moreover, 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 2B), which 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 [62], is less affected in PD than the lateral substantia nigra, which projects to dorsal striatum [31, 32].

Also, the structural and functional connectomes used here were derived from healthy individuals, as typically done [9, 10, 19, 60]. However, disease-related alterations to cell integrity should eventually affect network function. Connectomics data in PD are starting to become available and it will be interesting to incorporate these time-varying effects into our agent-based model.

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. (However, see S8 Fig for evidence that the model is robust to these parameter choices.)

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

Materials and methods

Human brain parcellation

We used a brain parcellation generated by atlas-based segmentation [28]. 68 cortical parcels were defined using curvature-based information [27], 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 [63]. Finally, substantia nigra was added to the atlas using the location provided in the ATAG atlas (https://www.nitrc.org/projects/atag) [29]. 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 [64]. Moreover, regional gene expression was mostly available only for the left hemisphere (see *Regional gene expression*).

PPMI patient data and image processing

PPMI is an open-access comprehensive observational clinical study [26], 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) [19], a method to detect local changes in tissue density. DBM was performed using the minc-toolkit available at https://github.com/BIC-MNI/minc-toolkit-v2.

After denoising [65], inhomogeneity correction [66], and linear intensity scaling, individual MRI images are registered non-linearly to the MNI152-2009c template [67], 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. Considering that in the denoising stage a non-local smoothing filter was applied to the T1 images, we decided to exclude substantia nigra in the estimation of atrophy as it is too small in size compared to the smoothing parameter and the deformation map may therefore not

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reflect the true level of tissue loss in such a small structure. Therefore, although substangia nigra is included in the spreading model (and plays a vital role), the model fit was assessed using only the 41 much larger cortical and subcortical regions (see [19] for more). The deformation maps can be found in https://neurovault.org/collections/860/.

Regional gene expression

Regional gene expression levels were derived from the six post-mortem brains included in the AHBA [68], 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) [69], 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 (S9 Fig).

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) [70] 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 [71].

Deterministic fiber tracking was conducted in native space using DSI studio (www.dsi-studio.labsolver.org) [22]. 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/) [33] 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

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

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 tissue loss (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 [56].

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. Therefore $(1 - \gamma_i^0)^{M_i}$ is the probability that a single normal agent is not infected by any of the M_i misfoled agents so that $\gamma_i = 1 - (1 - \gamma_i^0)^{M_i} = 1 - \mathrm{e}^{M_i \ln(1-\gamma_i^0)}$ denotes the probability of getting infected by at least one of the M_i misfolded agents in region i per unit time [11, 38]. 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

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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:

$$\Delta N_i = \alpha_i S_i \Delta t - (1-\mathrm{e}^{-\beta_i\Delta t})N_i - (\mathrm{e}^{-\beta_i\Delta t})(1-\mathrm{e}^{-\gamma_i^0 M_i\Delta t})N_i \tag{3}$$

$$\Delta M_i = (\mathrm{e}^{-\beta_i\Delta t})(1-\mathrm{e}^{-\gamma_i^0 M_i\Delta t})N_i - (1-\mathrm{e}^{-\beta_i\Delta t})M_i \tag{4}$$

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 S1 Text). 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:

$$P_{\text{region } i\to\text{region } i} = \rho_i \tag{5}$$

$$P_{\text{region } i\to\text{edge } (i,j)} = (1-\rho_i)\frac{w_{ij}}{\sum_j w_{ij}} \tag{6}$$

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 S9A Fig 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

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

Another important question to consider is whether the model fit we observed arises only from our particular choice of synthesis/clearance and propagation rate. Specifying the synthesis/clearance rate as a value between 0 and 1 transformed from the gene expression z-scores, we have simplified the complex relationship between transcriptions and the actual function of the gene; likewise, setting the probability of exiting an edges simply to the reciprocal of edge length, we have implicitly specified the relative scale and regional synthesis/clearance and propagation processes. It is possible that theses assumptions implicitly imposed on the model might not be able to reflect the actual spreading process. However, we found that the model yielded robust results as long as the relative magnitude of variations in regional gene expression z-scores is preserved in synthesis/clearance rate (S10A Fig). Moreover, varying the scale of propagation rate with respect to the synthesis/clearance process within certain range has little effect on model fit as well (S10B Fig).

Accrual of neuronal tissue loss (atrophy). We model neuronal tissue loss as the result of two processes: direct toxicity from accumulation of native misfolded α-synuclein and deafferentation (reduction in neuronal 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 S10B Fig 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). Code of the model and relevant data can be found in https://github.com/yingqiuz/SIR_simulator.

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 and right-left 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

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

Acknowledgements

We thank Ross Hammond for fruitful discussions on the idea of applying agent-based models to neurodegeneration.

References

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

2. Guo JL, Lee VM. Cell-to-cell transmission of pathogenic proteins in neurodegenerative diseases. Nature medicine. 2014;20(2):130.

3. Polymenidou M, Cleveland DW. Prion-like spread of protein aggregates in neurodegeneration. Journal of Experimental Medicine. 2012;209(5):889–893.

4. Jucker M, Walker LC. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. Nature. 2013;501(7465):45.

5. Brettschneider J, Del Tredici K, Lee VMY, Trojanowski JQ. Spreading of pathology in neurodegenerative diseases: a focus on human studies. Nature Reviews Neuroscience. 2015;16(2):109.

6. Walsh DM, Selkoe DJ. A critical appraisal of the pathogenic protein spread hypothesis of neurodegeneration. Nature Reviews Neuroscience. 2016;17(4):251.

7. Surmeier DJ, Obeso JA, Halliday GM. Selective neuronal vulnerability in Parkinson disease. Nature Reviews Neuroscience. 2017;18(2):101.

8. Kordower JH, Chu Y, Hauser RA, Freeman TB, Olanow CW. Lewy body–like pathology in long-term embryonic nigral transplants in Parkinson's disease. Nature medicine. 2008;14(5):504.

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

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

10. Iturria-Medina Y, Sotero RC, Toussaint PJ, Evans AC, Initiative ADN, et al. Epidemic spreading model to characterize misfolded proteins propagation in aging and associated neurodegenerative disorders. PLoS computational biology. 2014;10(11):e1003956.

11. Frias-Martinez E, Williamson G, Frias-Martinez V. 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. IEEE; 2011. p. 57–64.

12. Luk KC, Kehm V, Carroll J, Zhang B, O'Brien P, Trojanowski JQ, et al. Pathological α-synuclein transmission initiates Parkinson-like neurodegeneration in nontransgenic mice. Science. 2012;338(6109):949–953.

13. Mougenot AL, Nicot S, Bencsik A, Morignat E, Verchère J, Lakhdar L, et al. Prion-like acceleration of a synucleinopathy in a transgenic mouse model. Neurobiology of aging. 2012;33(9):2225–2228.

14. Goedert M, Spillantini MG, Del Tredici K, Braak H. 100 years of Lewy pathology. Nature Reviews Neurology. 2013;9(1):13.

15. Masuda-Suzukake M, Nonaka T, Hosokawa M, Oikawa T, Arai T, Akiyama H, et al. Prion-like spreading of pathological α-synuclein in brain. Brain. 2013;136(4):1128–1138.

16. Peelaerts W, Bousset L, Van der Perren A, Moskalyuk A, Pulizzi R, Giugliano M, et al. α-Synuclein strains cause distinct synucleinopathies after local and systemic administration. Nature. 2015;522(7556):340.

17. Rey NL, Steiner JA, Maroof N, Luk KC, Madaj Z, Trojanowski JQ, et al. Widespread transneuronal propagation of α-synucleinopathy triggered in olfactory bulb mimics prodromal Parkinson's disease. Journal of Experimental Medicine. 2016; p. jem–20160368.

18. Brundin P, Kordower JH. Neuropathology in transplants in Parkinson's disease: implications for disease pathogenesis and the future of cell therapy. In: Progress in brain research. vol. 200. Elsevier; 2012. p. 221–241.

19. Zeighami Y, Ulla M, Iturria-Medina Y, Dadar M, Zhang Y, Larcher KMH, et al. Network structure of brain atrophy in de novo Parkinson's disease. Elife. 2015;4.

20. Yau Y, Zeighami Y, Baker T, Larcher K, Vainik U, Dadar M, et al. Network connectivity determines cortical thinning in early Parkinson's disease progression. Nature communications. 2018;9(1):12.

21. Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K, et al. The WU-Minn human connectome project: an overview. Neuroimage. 2013;80:62–79.

22. Yeh FC, Verstynen TD, Wang Y, Fernández-Miranda JC, Tseng WYI. Deterministic diffusion fiber tracking improved by quantitative anisotropy. PloS one. 2013;8(11):e80713.

23. Mišić B, Betzel RF, Nematzadeh A, Goni J, Griffa A, Hagmann P, et al. Cooperative and competitive spreading dynamics on the human connectome. Neuron. 2015;86(6):1518–1529.

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

24. Mišić B, Betzel RF, Griffa A, de Reus MA, He Y, Zuo XN, et al. Network-Based Asymmetry of the Human Auditory System. Cerebral Cortex. 2018;28(7):2655–2664.

25. Betzel RF, Griffa A, Hagmann P, Mišić B. Distance-dependent consensus thresholds for generating group-representative structural brain networks. Network neuroscience. 2019;3(2):475–496.

26. Marek K, Jennings D, Lasch S, Siderowf A, Tanner C, Simuni T, et al. The parkinson progression marker initiative (PPMI). Progress in neurobiology. 2011;95(4):629–635.

27. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006;31(3):968–980.

28. Cammoun L, Gigandet X, Meskaldji D, Thiran JP, Sporns O, Do KQ, et al. Mapping the human connectome at multiple scales with diffusion spectrum MRI. Journal of neuroscience methods. 2012;203(2):386–397.

29. Keuken MC, Bazin PL, Crown L, Hootsmans J, Laufer A, Müller-Axt C, et al. Quantifying inter-individual anatomical variability in the subcortex using 7T structural MRI. NeuroImage. 2014;94:40–46.

30. Rey NL, George S, Steiner JA, Madaj Z, Luk KC, Trojanowski JQ, et al. Spread of aggregates after olfactory bulb injection of α-synuclein fibrils is associated with early neuronal loss and is reduced long term. Acta neuropathologica. 2018;135(1):65–83.

31. Braak H, Del Tredici K, Rüb U, De Vos RA, Steur ENJ, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiology of aging. 2003;24(2):197–211.

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

33. Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. Neuroimage. 2012;62(2):782–790.

34. Crossley NA, Mechelli A, Scott J, Carletti F, Fox PT, McGuire P, et al. The hubs of the human connectome are generally implicated in the anatomy of brain disorders. Brain. 2014;137(8):2382–2395.

35. Luna E, Decker SC, Riddle DM, Caputo A, Zhang B, Cole T, et al. Differential α-synuclein expression contributes to selective vulnerability of hippocampal neuron subpopulations to fibril-induced toxicity. Acta neuropathologica. 2018; p. 1–21.

36. Spillantini MG, Crowther RA, Jakes R, Hasegawa M, Goedert M. α-Synuclein in filamentous inclusions of Lewy bodies from Parkinson's disease and dementia with Lewy bodies. Proceedings of the National Academy of Sciences. 1998;95(11):6469–6473.

37. Damier P, Hirsch E, Agid Y, Graybiel A. The substantia nigra of the human brain: II. Patterns of loss of dopamine-containing neurons in Parkinson's disease. Brain. 1999;122(8):1437–1448.

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

38. Newman M. Networks: an introduction. Oxford university press; 2010.

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

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

41. Roberts JA, Perry A, Lord AR, Roberts G, Mitchell PB, Smith RE, et al. The contribution of geometry to the human connectome. Neuroimage. 2016;124:379–393.

42. Seguin C, van den Heuvel MP, Zalesky A. Navigation of brain networks. Proceedings of the National Academy of Sciences. 2018; p. 201801351.

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

44. Fereshtehnejad SM, Zeighami Y, Dagher A, Postuma RB. Clinical criteria for subtyping Parkinson's disease: biomarkers and longitudinal progression. Brain. 2017;140(7):1959–1976.

45. Brockmann D, Helbing D. The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. Science. 2013;342(6164):1337–1342. doi:10.1126/science.1245200.

46. Weickenmeier J, Kuhl E, Goriely A. Multiphysics of Prionlike Diseases: Progression and Atrophy. Phys Rev Lett. 2018;121:158101. doi:10.1103/PhysRevLett.121.158101.

47. Watts DJ, Strogatz SH. Collective dynamics of 'small-world'networks. nature. 1998;393(6684):440.

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

49. Hilgetag CC, Kaiser M. Clustered organization of cortical connectivity. Neuroinformatics. 2004;2(3):353–360.

50. Sporns O, Betzel RF. Modular brain networks. Annual review of psychology. 2016;67:613–640.

51. Nematzadeh A, Ferrara E, Flammini A, Ahn YY. Optimal network modularity for information diffusion. Physical review letters. 2014;113(8):088701.

52. van den Heuvel MP, Kahn RS, Goñi J, Sporns O. High-cost, high-capacity backbone for global brain communication. Proceedings of the National Academy of Sciences. 2012;109(28):11372–11377.

53. Volpicelli-Daley LA, Luk KC, Patel TP, Tanik SA, Riddle DM, Stieber A, et al. Exogenous α-synuclein fibrils induce Lewy body pathology leading to synaptic dysfunction and neuron death. Neuron. 2011;72(1):57–71.

54. Alcalay RN, Levy OA, Waters CH, Fahn S, Ford B, Kuo SH, et al. Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations. Brain. 2015;138(9):2648–2658.

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

55. Aharon-Peretz J, Rosenbaum H, Gershoni-Baruch R. Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. New England Journal of Medicine. 2004;351(19):1972–1977.

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

57. Singleton A, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, et al. α-Synuclein locus triplication causes Parkinson's disease. Science. 2003;302(5646):841–841.

58. Chartier-Harlin MC, Kachergus J, Roumier C, Mouroux V, Douay X, Lincoln S, et al. α-synuclein locus duplication as a cause of familial Parkinson's disease. The Lancet. 2004;364(9440):1167–1169.

59. Ibanez P, Bonnet A, Debarges B, Lohmann E, Tison F, Agid Y, et al. Causal relation between α-synuclein locus duplication as a cause of familial Parkinson's disease. The Lancet. 2004;364(9440):1169–1171.

60. Seeley WW, Crawford RK, Zhou J, Miller BL, Greicius MD. Neurodegenerative diseases target large-scale human brain networks. Neuron. 2009;62(1):42–52.

61. Michel PP, Hirsch EC, Hunot S. Understanding dopaminergic cell death pathways in Parkinson disease. Neuron. 2016;90(4):675–691.

62. Zhang Y, Larcher KMH, Misic B, Dagher A. Anatomical and functional organization of the human substantia nigra and its connections. Elife. 2017;6:e26653.

63. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 2002;33(3):341–355.

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

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

66. Sled JG, Zijdenbos AP, Evans AC. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE transactions on medical imaging. 1998;17(1):87–97.

67. Collins DL, Evans AC. Animal: validation and applications of nonlinear registration-based segmentation. International journal of pattern recognition and artificial intelligence. 1997;11(08):1271–1294.

68. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature. 2012;489(7416):391.

69. Gorgolewski K, Fox A, Chang L, Schäfer A, Arélin K, Burmann I, et al. Tight fitting genes: Finding relations between statistical maps and gene expression patterns. Organization for Human Brain Mapping Hamburg, Germany. 2014;.

70. Yeh FC, Wedeen VJ, Tseng WYI. Generalized q-Sampling Imaging. IEEE transactions on medical imaging. 2010;29(9):1626–1635.

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

71. Yeh FC, Tseng WYI. NTU-90: a high angular resolution brain atlas constructed by q-space diffeomorphic reconstruction. Neuroimage. 2011;58(1):91–99.

72. McCann H, Cartwright H, Halliday GM. Neuropathology of α-synuclein propagation and braak hypothesis. Movement Disorders. 2016;31(2):152–160.

73. Hof PR, Morrison JH. The aging brain: morphomolecular senescence of cortical circuits. Trends in neurosciences. 2004;27(10):607–613.

Supporting Information

S1 Text. Analysis of the fixed points.

S1 Table. A list of all the parameters or notations used in the model. 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 (S10 Fig).

S1 Fig. 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 (S4)) and N increases with M (M nullcline, orange, equation (S5)), 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). The code to generate the figure can be found in https://github.com/yingqiuz/SIR_simulator/tree/master/results/S1_Fig.ipynb.

S2 Fig. 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. The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/tree/master/results/S2_Fig.mat.

S3 Fig. Arrival time of misfolded α-synuclein in the model. (A) 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 (see [31]). (B) Arrival time of misfolded α-synuclein at each brain region. The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/tree/master/results/S3_Fig.mat.

S4 Fig. Model fit with atrophy estimated using fsl_anat. The computational model replicates empirical atrophy patterns estimated using an alternative deformation-based morphometry pipeline. Model fits are comparable between the *minctools* and *FSL*-estimated atrophy patterns. The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/tree/master/results/S4_Fig.mat.

bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S5 Fig. 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. The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/tree/master/results/S5_Fig.mat.

S6 Fig. Model fit at 65-region and 119-region resolution. The 42 regions used in the main manuscript were hierarchically partitioned into 65 regions and then a further 119 regions. Simulations were conducted on these two finer resolutions, and yielded comparable results to the model fit at 42-region resolution. (A) Spearman's correlation (blue curve) and Pearson's correlation (red curve) versus time using the 65-region parcellation. Black dot: peak position of the correlation coefficients. (B) The model has more predictive power than its constituent factors (as assessed by Spearman's correlation). C: Spearman's correlation (blue curve) and Pearson's correlation (red curve) versus time using the 119-region parcellation. Black dot: peak position of the correlation coefficients. D: The model has more predictive power than its constituent factors (as assessed by Spearman's correlation). The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/tree/master/results/S6_Fig.mat.

S7 Fig. Permutation tests for functional connectivity (FC). Increasing k (the influence of FC on α-synuclein transmission) first facilitates then degrades model fit. The red line indicates model fit using true FC values. For each k, rs-fMRI time series were re-assigned to construct null FC matrices. The null model fit declines monotonously as k increases (gray line). At smaller vaues of k, simulations based on real FC yield significantly higher model fit than the null settings as indicated by the 95% confidence interval (gray bar), while at larger k, real FC ceases to have advantage over null FC. The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/tree/master/results/S7_Fig.mat.

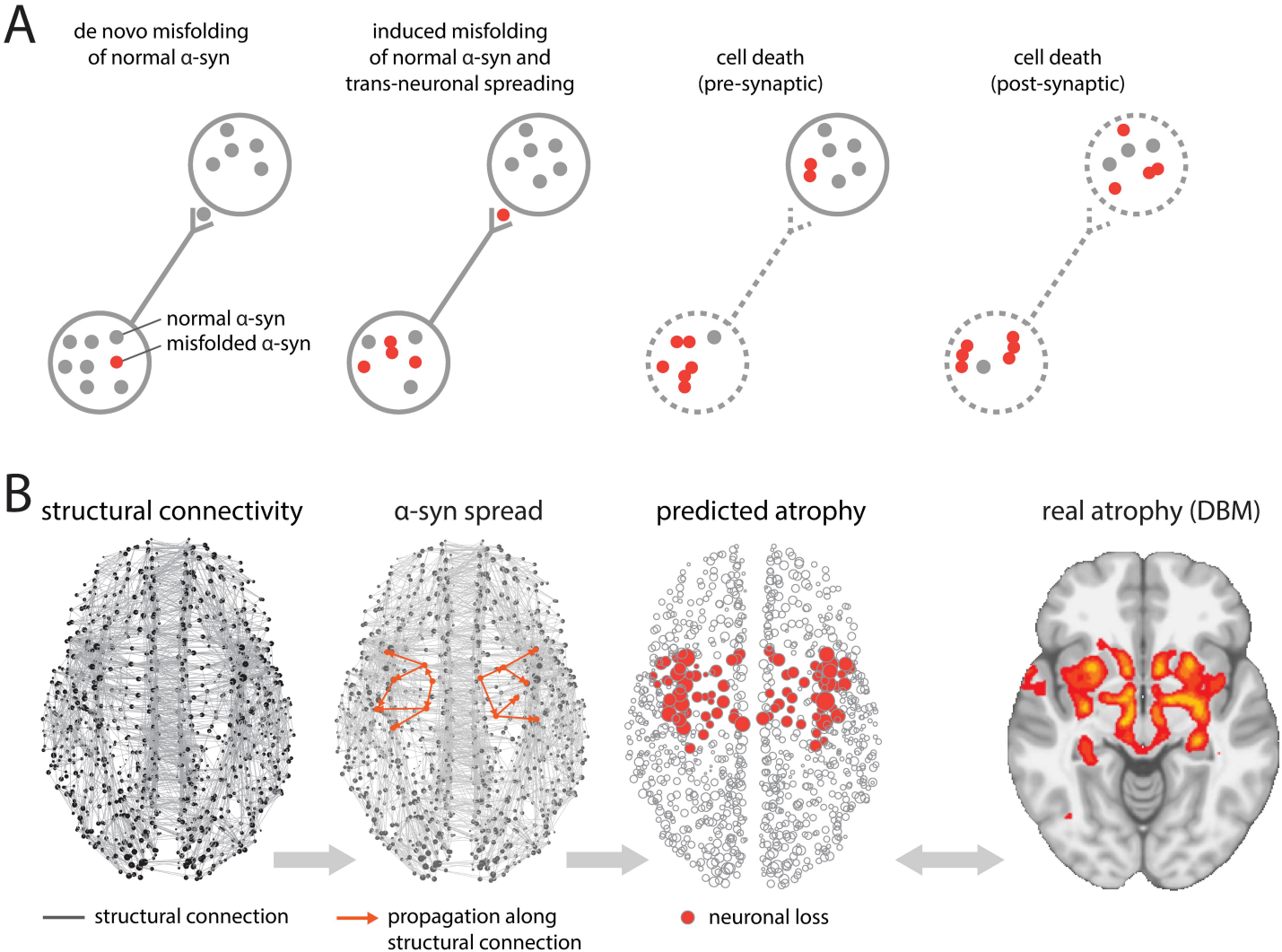
S8 Fig. Exploring other choices of synthesis/clearance and propagation rates. (A) In the main setting, we transformed the *SNCA/GBA* expression z-scores to synthesis/clearance rates using a standard normal cumulative distribution function (Fig. 2). To test other possibilities of the relation between gene expression and synthesis or clearance rate, we chose a set of commonly used functions that have domain of all real numbers and return values monotonically from 0 to 1. The temporal pattern of model fit is robust to the choice of transformation functions: the model yields similar results as long as the relative magnitudes of regional gene expressions are preserved. x_i is the gene expression z-score in region i and $f(x_i)$ is the transformation function. (B) We set the probability of exiting an edge (i, j) to the reciprocal of edge length $l_{i,j}$ in the main setting. However, it is possible that the protein agents propagate faster or slower than the regional synthesis/clearance process. To test sensitivity of the model to the rate of protein propagation, we introduced propagation speed v and set the probability thereof to $v/l_{i,j}$ such that varying v changes the relative scale of the propagation process vis-a-vis the regional synthesis/clearance process (e.g., increasing v suggests that the propagation process happens faster than the regional synthesis/clearance processes). We chose $v = 0.1, 1, 10$ (where $v = 1$ corresponds to the results in the main text) and found that the relative scale of the two processes has little effect on the model fit. https://github.com/yingqiuz/SIR_simulator/tree/master/results/S8_Fig.mat.

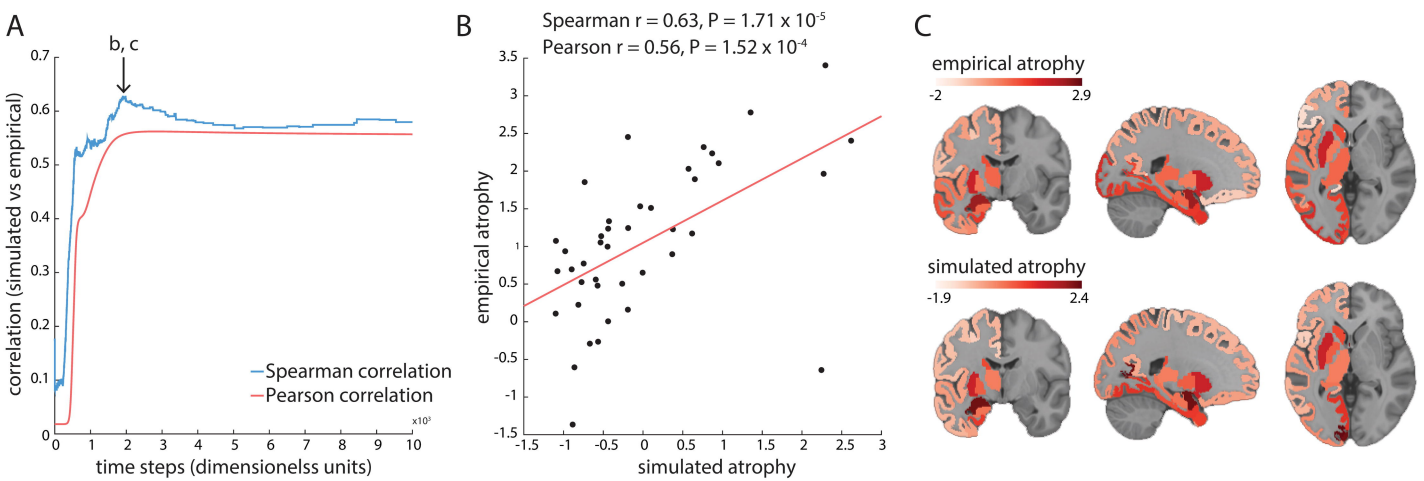
S9 Fig. *GBA* and *SNCA* expression. (A) Regional *GBA* expression. There are

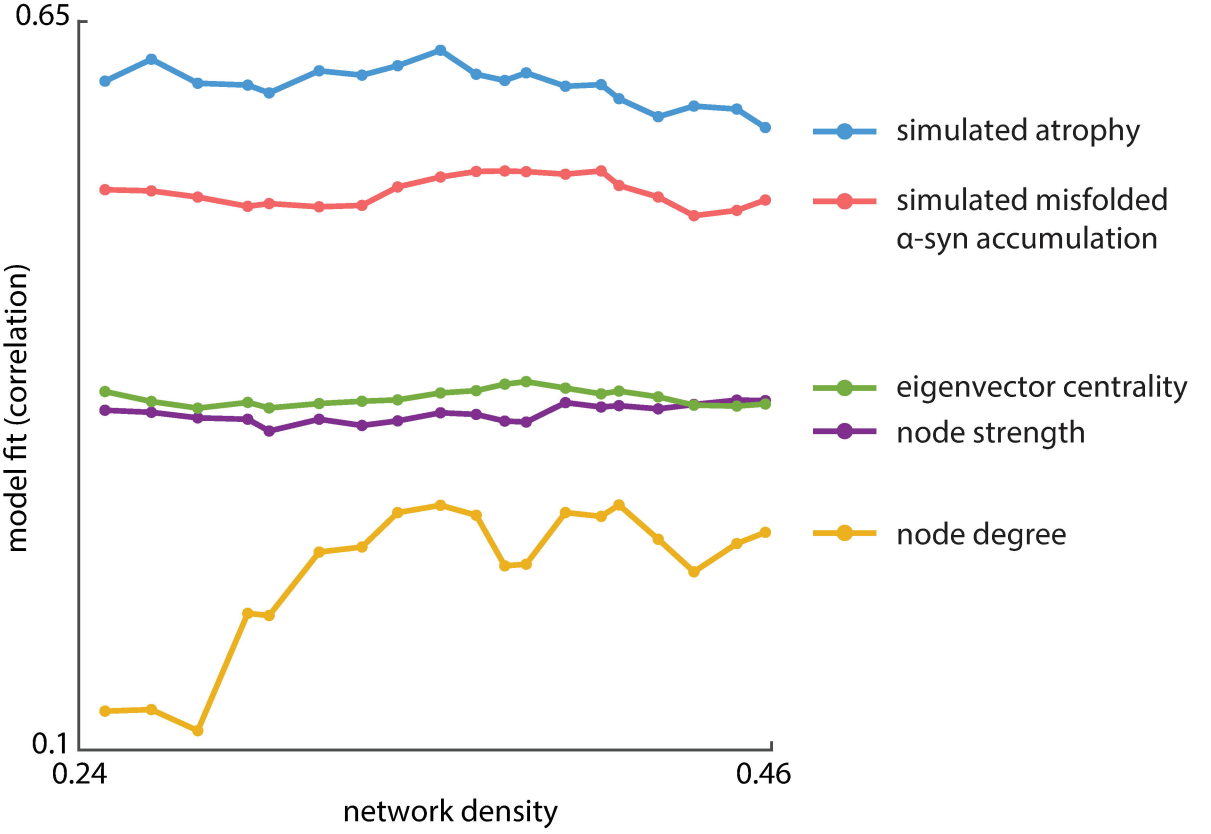
bioRxiv preprint doi: https://doi.org/10.1101/449199; this version posted November 1, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

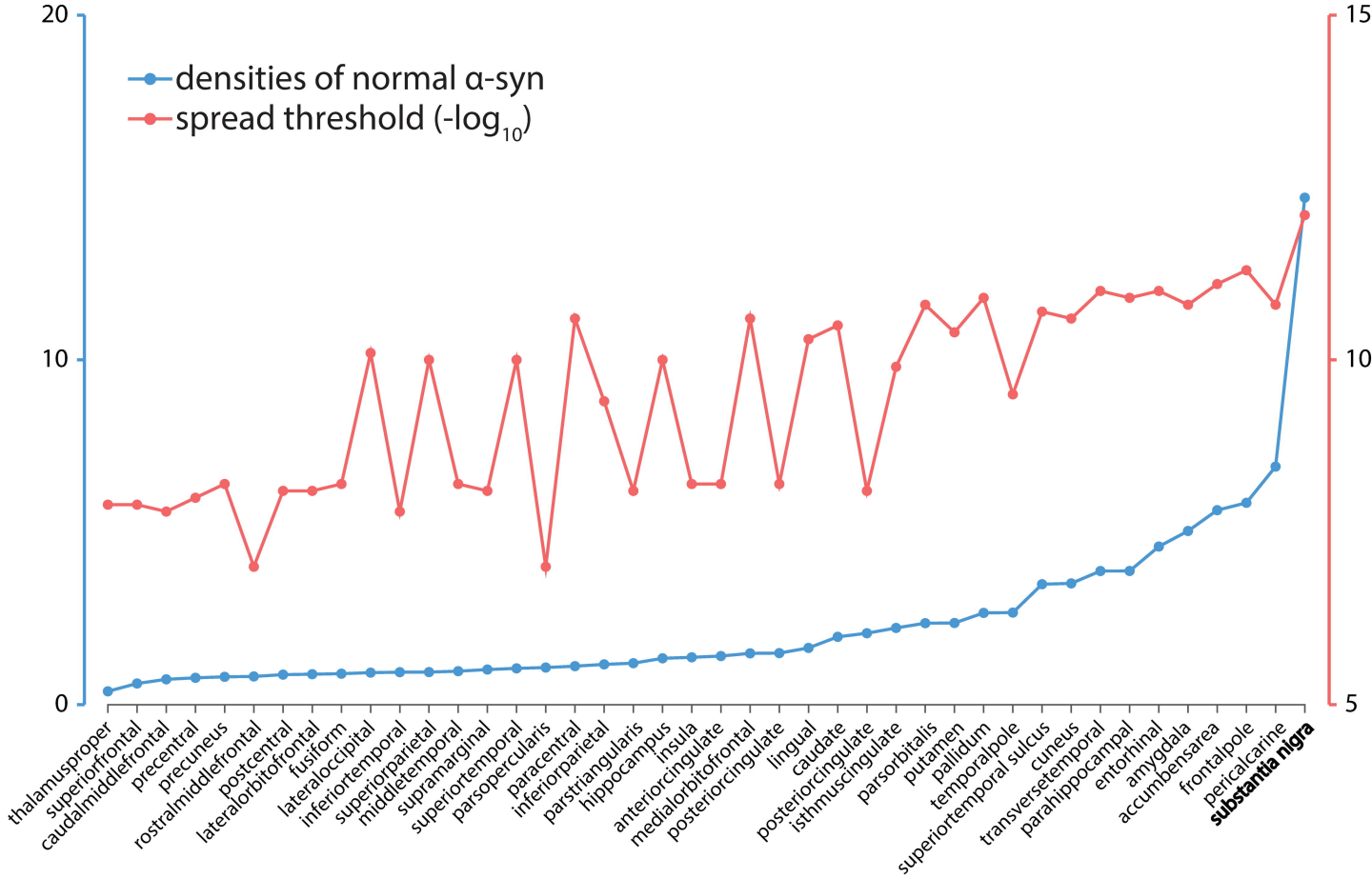
three probes for *GBA* (probe id: 1025372, 1025373, 1025374). Probes 1025372 and 1025373 were included to generate the group transcription profile. Probe 1025374 was excluded as it deviated too much from probe 1025372 (Pearson correlation=0.30), while the correlation between the other two probes is 0.79. (B) Regional *SNCA* expression. Probes 1020182, 1010655 were included to generate the group transcription map. Compared to *GBA* expression, *SNCA* is more homogeneous in cortical regions. The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/tree/master/results/S9_Fig.mat.

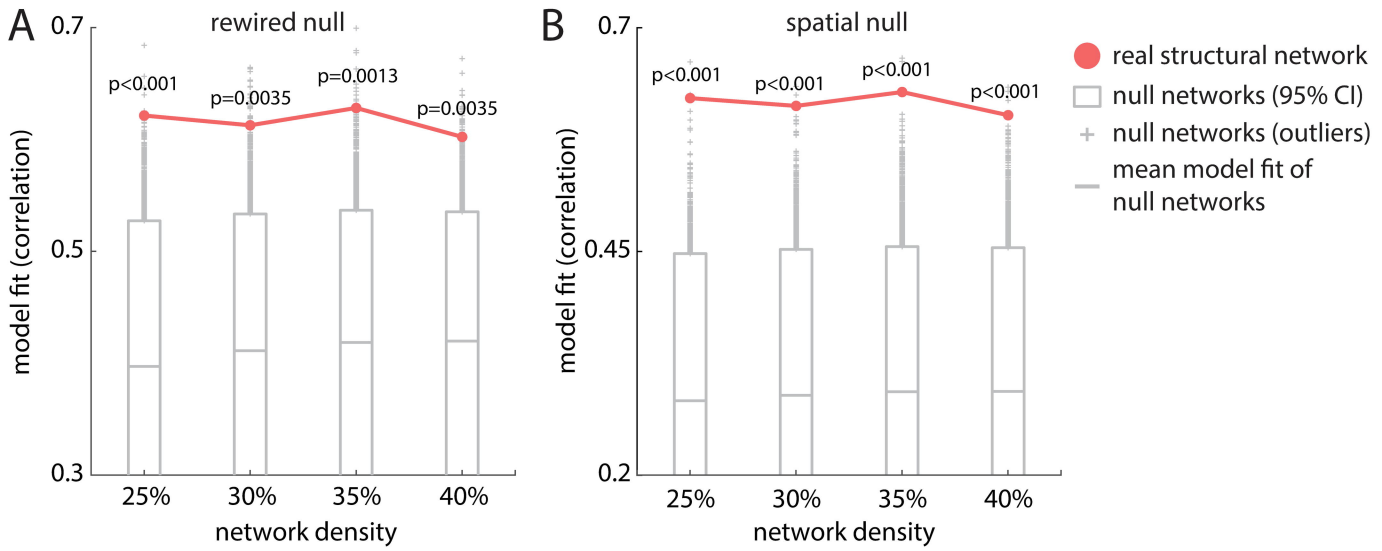
S10 Fig. Testing free parameters ρ_i, k_1, k_2. Model fit (Spearman's correlation) is robust to variations in ρ_i, k_1, k_2 (results shown at network density %35). (A) ρ_i controls the probability of remaining in region i while (1-ρ_i) is the probability of exiting region i per unit time. The main results 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. The underlying data can be found in https://github.com/yingqiuz/SIR_simulator/tree/master/results/S10_Fig.mat.

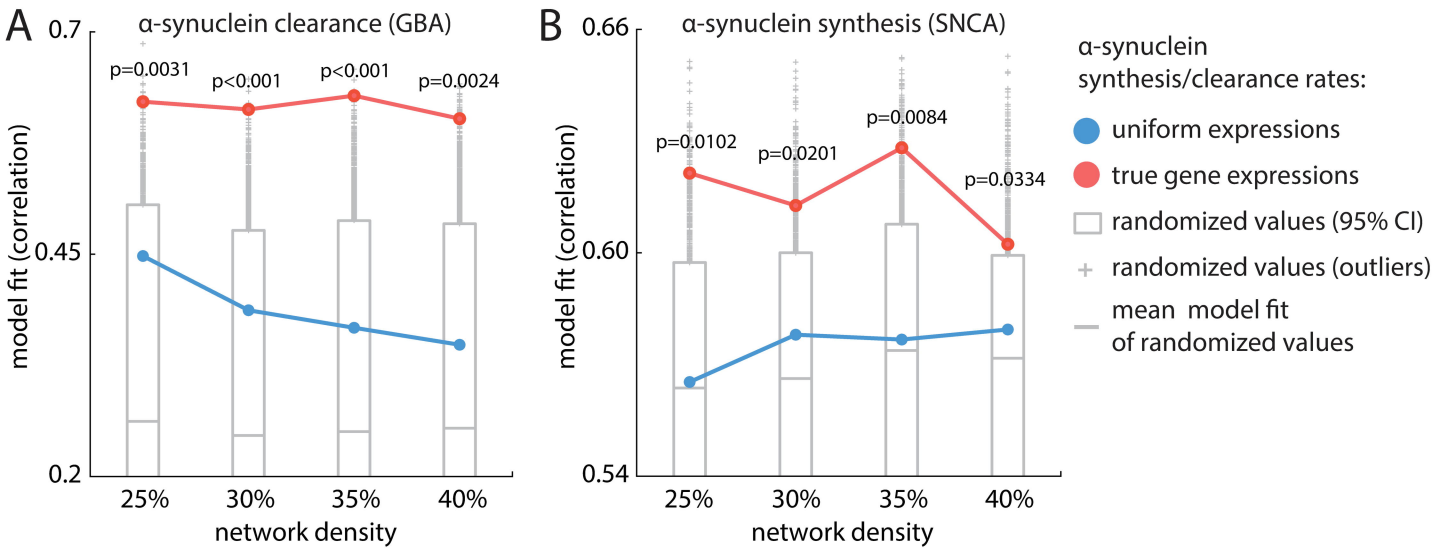

A
de novo misfolding of normal α-syn
induced misfolding of normal α-syn and trans-neuronal spreading
cell death (pre-synaptic)
cell death (post-synaptic)
normal α-syn
misfolded α-syn
B
structural connectivity
α-syn spread
predicted atrophy
real atrophy (DBM)
structural connection
propagation along structural connection
neuronal loss

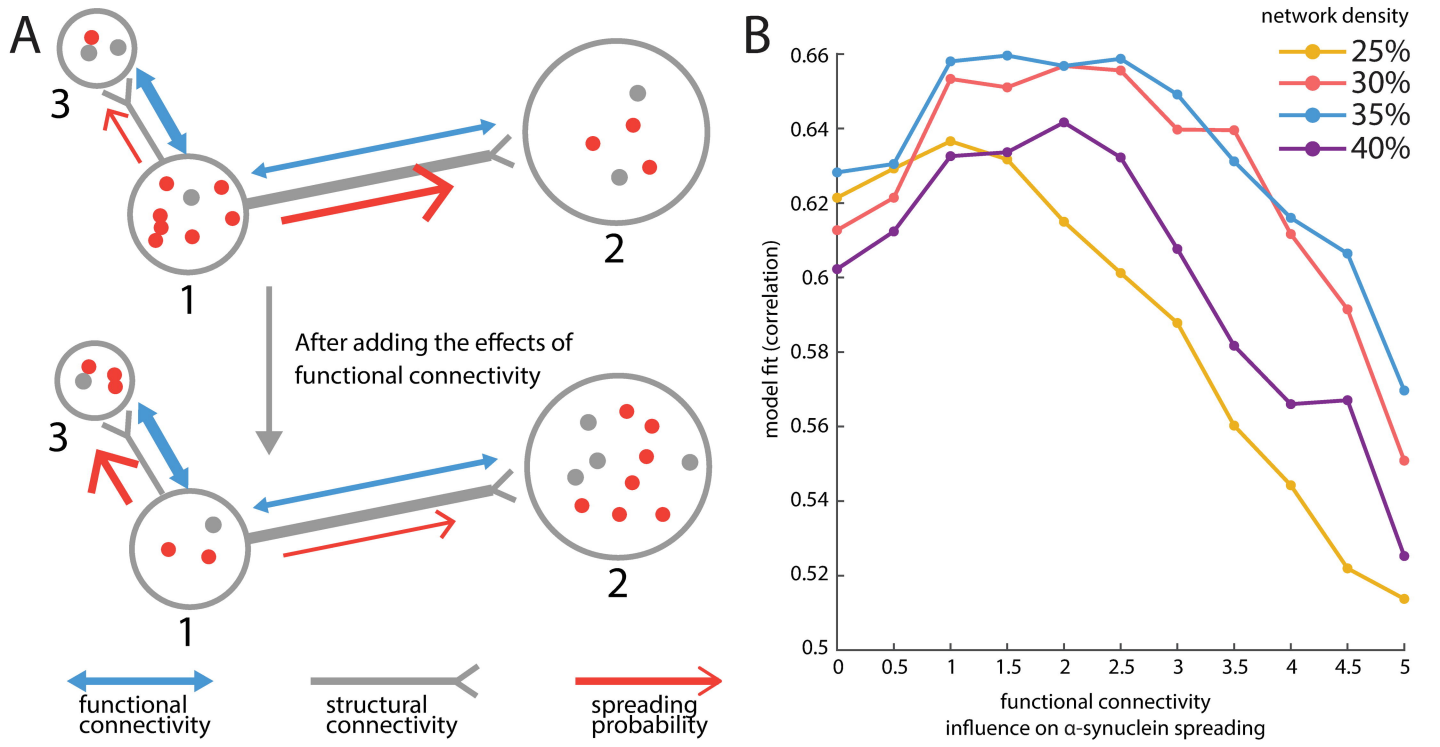

A
b, c
correlation (simulated vs empirical)
time steps (dimensionelss units)
Spearman correlation
Pearson correlation
B
Spearman r = 0.63, P = 1.71 x 10-5
Pearson r = 0.56, P = 1.52 x 10-4
empirical atrophy
simulated atrophy
C
empirical atrophy
-2
2.9
simulated atrophy
-1.9
2.4


0.65
0.1
model fit (correlation)
0.24
0.46
network density
simulated atrophy
simulated misfolded
α-syn accumulation
eigenvector centrality
node strength
node degree


20
10
0
15
10
5
densities of normal α-syn
spread threshold (-log10)
thalamusproper
superiorfrontal
caudalmiddlefrontal
precentral
precuneus
rostralmiddlefrontal
postcentral
lateralorbitofrontal
fusiform
lateraloccipital
inferiortemporal
superiorparietal
middletemporal
supramarginal
superiortemporal
parsopercularis
paracentral
inferiorparietal
parstriangularis
hippocampus
insula
anteriorcingulate
medialorbitofrontal
posteriorcingulate
lingual
caudate
posteriorcingulate
isthmuscingulate
parsorbitalis
putamen
pallidum
temporalpole
superiortemporal sulcus
cuneus
transversetemporal
parahippocampal
entorhinal
amygdala
accumbensarea
frontalpole
pericalcarine
substantia nigra


A
rewired null
0.7
0.5
0.3
model fit (correlation)
p<0.001
p=0.0035
p=0.0013
p=0.0035
25%
30%
35%
40%
network density
B
spatial null
0.7
0.45
0.2
model fit (correlation)
p<0.001
p<0.001
p<0.001
p<0.001
25%
30%
35%
40%
network density
real structural network
null networks (95% CI)
null networks (outliers)
mean model fit of null networks


A
α-synuclein clearance (GBA)
0.7
0.45
0.2
model fit (correlation)
p=0.0031
p<0.001
p<0.001
p=0.0024
25%
30%
35%
40%
network density
B
α-synuclein synthesis (SNCA)
0.66
0.60
0.54
model fit (correlation)
p=0.0102
p=0.0201
p=0.0084
p=0.0334
25%
30%
35%
40%
network density
α-synuclein synthesis/clearance rates:
uniform expressions
true gene expressions
randomized values (95% CI)
randomized values (outliers)
mean model fit of randomized values


A
3
1
2
After adding the effects of functional connectivity
3
1
2
functional connectivity
structural connectivity
spreading probability
B
network density
25%
30%
35%
40%
model fit (correlation)
0.66
0.64
0.62
0.6
0.58
0.56
0.54
0.52
0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
functional connectivity influence on α-synuclein spreading