Differences in topological progression profile among neurodegenerative diseases from imaging data

Sara Garbarino1,2,*, Marco Lorenzi2, Neil P. Oxtoby1, Elisabeth J. Vinke3, Razvan V. Marinescu1, Arman Eshaghi4, M. Arfan Ikram5,7, Wiro J. Niessen7, Olga Ciccarelli4, Frederik Barkhof1,5, Jonathan M. Schott6, Meike W. Vernooij3,7, Daniel C. Alexander1, for the Alzheimer’s Disease Neuroimaging Initiative

1 Centre for Medical Image Computing, University College London, United Kingdom
2 Université Côte d’Azur, Inria, France
3 Department of Epidemiology, Erasmus Medical Center, the Netherlands
4 Institute of Neurology, University College London, United Kingdom
5 Department of Radiology and Nuclear medicine, VUmc, the Netherlands
6 Dementia Research Centre, Institute of Neurology, University College London, United Kingdom
7 Department of Radiology and Nuclear medicine, Erasmus MC, the Netherlands

*Corresponding author: Sara Garbarino; email: sara.garbarino@inria.fr

Abstract

The spatial distribution of atrophy in neurodegenerative diseases suggests that brain connectivity mediates disease propagation. Different descriptors of the connectivity graph potentially relate to different underlying mechanisms of propagation. Previous approaches for evaluating the influence of connectivity on neurodegeneration consider each descriptor in isolation and match predictions against late-stage atrophy patterns. We introduce the notion of a topological profile — a characteristic combination of topological descriptors that best describes the propagation of pathology in a particular disease. By drawing on recent advances in disease progression modeling, we estimate topological profiles from the full course of pathology accumulation, at both cohort and individual levels. Experimental results comparing topological profiles for Alzheimer’s disease, multiple sclerosis and normal ageing show that topological profiles explain the observed data better than single descriptors. Within each condition, most individual profiles cluster around the cohort-level profile, and individuals whose profiles align more closely with other cohort-level profiles show features of that cohort. The cohort-level profiles suggest new insights into the biological mechanisms underlying pathology propagation in each disease.

Introduction

Evidence from neuroimaging suggests that the progression of brain changes in neurodegenerative diseases may be mediated by brain connectivity. For example, atrophy patterns observed with MRI suggest that different brain regions are systematically and selectively vulnerable to different neurodegenerative diseases, and that these atrophy patterns closely match known connectivity networks1–5. The literature includes wide
debate on the potential mechanisms underlying pathogenic protein propagation, on the biology of protein aggregation and propagation, and of selective vulnerability of neurons in neurodegenerative disease; see 6–8 for recent reviews. Uncovering the precise relationship between the topology of brain network connectivity and the pattern of pathology may provide clues to the underlying mechanisms of propagation. Indeed, in Alzheimer’s disease (AD), Zhou et al 3 attempt to explain the association between patterns of brain atrophy and connectivity in terms of different topological descriptors: 1) centrality – the most active regions or ‘hubs’ are most vulnerable 9,10; 2) segregation – opposite to 1), the most isolated nodes are most vulnerable 11,12; 3) network proximity – regions connected to disease epicenters are most vulnerable 13–15; 4) cortical proximity – regions spatially-adjacent to disease epicenters are most vulnerable. The authors relate these descriptors to distinct mechanisms of pathology propagation: (1) centrality represents nodal stress, (2) segregation, trophic failure, (3) network proximity, trans-neuronal spread, and (4) cortical proximity represents unguided diffusive propagation. Thus comparing patterns of pathology predicted by these different descriptors with those observed in patient cohorts gives clues to which corresponding mechanisms are at play. Moreover, as reviews 6–8 highlight, the plausibility of those mechanisms extend similarly to the wider range of neurodegenerative conditions.

Previous studies 1,2,4,5,10,16–19 using functional or structural networks have focused on evaluating the ability of each individual topological descriptor above to explain observed patient data with the goal of identifying the single most likely mechanism of disease propagation. More generally, a variety of mathematical models based on dynamical systems modeling 20–26 have been proposed for describing the temporal progression of pathology mediated by brain networks. They mostly enforce the trans-neuronal spread (or “prion-like”) hypothesis of propagation via network proximity. These approaches have important methodological limitations, which should prompt caution in their interpretation. Two key limitations are: i) the approaches do not consider the possibility of multiple concurrent spreading mechanisms; ii) the evaluation of topological descriptors is performed using only cross-sectional data assumed to represent end-stage atrophy patterns.

Regarding limitation (i), multiple distinct mechanisms are likely to contribute in diseases where multiple proteinopathies are at play (such as amyloid and tau in AD; see for example 22,27,28); in diseases exhibiting highly-variable atrophy patterns (as in multiple sclerosis (MS); see 29); or in individual cases where multiple pathologies co-exist, as for instance tau accumulation in Parkinson’s disease 30,31, or alpha-synuclein or TDP-43 in AD 32–35. Also, associations between vascular factors and neurodegenerative dementias such as AD are common 36,37, which suggests contributions from multiple underlying mechanisms to the observed pattern of pathology accrual 35. Further, in MS, retrograde neurodegeneration secondary to focal damage from remote lesions, iron accumulation in the deep gray matter and “virtual hypoxia” in the hub regions could all potentially combine and contribute to explain observed neurodegeneration 38.

Regarding limitation (ii), considering only end-stage pathology severely limits sensitivity — like trying to guess the plot of a movie after watching only the final scene 39. The emergence of data-driven disease progression models (DPMs) 39–54 provides an opportunity to address limitation (ii) by using the full trajectory of pathology accumulation to evaluate the influence of topological descriptors. These techniques estimate the long-term temporal pattern of disease progression directly from cross-sectional or short-term longitudinal data sets on the assumption of some degree of commonality of progression over a patient cohort. In contrast to brute-force machine-learning approaches 56–58 to predict progression, DPMs reveal the trajectory of temporal evolution of multiple biomarkers on a common timeline, which can, in turn, provide an extra dimension for identifying informative topological descriptors over and above using late-stage information alone.

In this paper we set out to ameliorate the aforementioned limitations in network-based models of neurodegenerative disease by revealing the combinations of topological descriptors that best explain the temporal evolution of pathology. We introduce a new method that uses the Gaussian Process (GP)
Progression Model of \(^{59}\) to assess candidate descriptors against the full time course of the disease, rather than just late-stage pathology. Moreover, the method identifies a characteristic combination for each disease that defines a novel disease-specific topological profile. We extend this concept further to find personalized topological profiles for each individual and demonstrate the consistency of individual profiles with the corresponding cohort profile as well as characterize within-disease variability of individual topological profiles. We focus on patterns and trajectories of atrophy accumulation informed by structural MRI and use three distinct data sets across the spectrum of neurodegeneration (Alzheimer’s disease – AD, primary progressive multiple sclerosis – PPMS, and broadly-healthy ageing represented by data collected from community-dwelling ageing individuals – HA). We show that: a) a combination of topological descriptors consistently explains the data better than the best single descriptor; b) the profiles differ substantially among conditions; c) individual profiles cluster around corresponding cohort-level profiles, but only when the profile is estimated from the full time course rather than end-stage only; and d) positioning of individual profiles with respect to each cohort profile is associated with relevant clinical features, thereby potentially benefiting early diagnosis and stratification.

## Results

Figure 1 gives an overview of the methods used to compute topological profiles at both cohort and individual levels: full details are provided in the Methods section. In this section we show first that the GP Progression Model estimates trajectories of atrophy evolution that reflect observed atrophy patterns in each cohort (AD, PPMS and HA). Secondly, we show that the topological profiles derived from such atrophy evolution patterns are distinct for each condition and explain observations better than profiles estimated when using end-stage data only, and better than single topological descriptors. Further, we demonstrate that most individual profiles reflect their corresponding cohort profile (and neurological condition), and that those that do not (“outliers”) show clinical characteristics of the topologically-nearest cohort.

Temporal patterns estimated by the GP Progression Model confirm observed atrophy progression patterns

Figure 2, top row, shows the spatio-temporal evolution of the atrophy patterns in each cohort (AD, PPMS and HA) estimated by the GP Progression Model over 41 bilateral regions of interest, obtained by symmetrizing the 82 anatomical regions from the segmentation procedure in \(^{60}\) (see Methods: Data description). Each panel shows four temporal stages sampled at uniform intervals according to the estimated disease time of the GP Progression Model (see Methods: GP Progression Model). Each topological profile specifies the rate of pathology accumulation in each area. So, while the rates are time-independent, they vary spatially so that the pattern itself is time dependent. Figure 2–figure supplement 1 shows a higher temporal resolution visualization of the same progressions. The regional trajectories and the individual time parameters estimated by the model are shown in Figure 2-figure supplement 2, Figure 2-figure supplement 3 and Figure 2-figure supplement 4.

In the AD cohort, the first regions to show atrophy are the superior temporal region and the hippocampus, followed by the amygdala, the remaining temporal regions, the insular and the supramarginal regions, and then the precentral and postcentral regions and the posterior lobe. The estimated progression gradually involves the occipital lobe, the middle frontal region, and finally the remaining subcortical areas, with thalamus and caudate last, which matches well-known atrophy progression patterns observed in AD from post-mortem histology \(^{61}\) and in-vivo disease progression models \(^{41,44,55,59}\). In PPMS the progression first involves some subcortical areas (caudate, thalamus, pallidum), followed by the superior parietal region, the remaining subcortical areas (amygdala, putamen, nucleus accumbens), few frontal regions (mostly middle- and orbito-frontal), the precentral, and then the occipital and temporal lobes, which become abnormal later in the progression. The dynamics of progression agree with recent results \(^{42,62}\), obtained using a different DPM. In the HA cohort, we observe early involvement of the insula, the superior and middle temporal lobes, the middle frontal, and the putamen. Subsequently, the amygdala, hippocampus and nucleus
accumbens are affected, the inferior temporal and more frontal regions, followed by the parietal lobe and the cingulate. This agrees with other studies of volume loss in normal aging.63,64,65

Distinct topological profiles for each neurological condition

The second row of Figure 2 shows the progression corresponding to the topological profiles that best match the GP Progression Model, for each cohort. The topological profiles are sparse linear combinations of nine network metrics, each representing one of five topological descriptors. They are the four described in – centrality, segregation, network proximity and cortical proximity – and the constant progression descriptor, quantifying the extent to which the rate of atrophy remains constant throughout the progression (see Methods: Network metrics). Table 1 shows the weight of each metric in the profile for each condition that best explains the corresponding GP disease progression in Figure 2 top row. Weights with value below 0.10 are set to zero in the final sparse linear combination defining the topological profiles; those weights are shown in bold in Table 1.

Table 1 shows that the topological profiles for AD, PPMS and HA differ substantially. Specifically, AD shows a concurrency of centrality metrics (40%) and network proximity (60%); PPMS shows a more complex profile with presence of centrality (35%), segregation (45%) and cortical proximity (20%); HA matches a combination of centrality (25%), cortical proximity (60%) and constant progression (20%).

Credible intervals in the topological profile weights, shown in parentheses in Table 1, reflect the credible intervals explicitly estimated by the GP Progression Model. Variability of the topological profile parameters under bootstrapping with 100 samples are shown in brackets. The reported p-values and effect sizes (in braces) are relative to the null hypothesis of \( \beta = 0 \) aside from the term associated with the constant progression, computed via permutation testing and Bonferroni-corrected for multiple comparison across the set of network metrics. All p-values were found <0.01 apart from the inverse degree for AD, for which \( p=0.048 \). Supplementary Information (Table S1) shows the topological profiles for all 82 brain regions; results remain consistent. Supplementary Information (Table S2) shows the topological profiles for two subsets of the HA cohort, age-matched with the AD and PPMS cohorts, as compared to topological profiles for the whole HA cohort; results remain consistent.

| Topological descriptor | Network metrics | AD               | PPMS             | HA               |
|------------------------|-----------------|------------------|------------------|------------------|
|                        | Betweenness     | **0.21(0.17)**   | **0.10(0.08)**   | 0.09(0.06)       |
|                        | Centrality      | [0.22(0.18)]     | [0.11(0.08)]     | [0.07(0.05)]     |
|                        |                 | {0.01(2.85)}     | {0.01(2.72)}     | {0.01(3.16)}     |
|                        | Closeness       | 0.01(0.02)       | **0.11(0.11)**   | 0.03(0.05)       |
|                        | Centrality      | [0.04(0.04)]     | [0.12(0.12)]     | [0.04(0.04)]     |
|                        |                 | {0.01(2.90)}     | {0.01(2.34)}     | {0.01(1.79)}     |
|                        | Weighted degree | 0.03(0.03)       | 0.07(0.05)       | **0.13(0.09)**   |
|                        |                 | [0.02(0.02)]     | [0.07(0.05)]     | [0.11(0.07)]     |
|                        |                 | {0.01(3.10)}     | {0.01(2.96)}     | {0.01(2.78)}     |
|                        | Clustering      | **0.19(0.11)**   | **0.14(0.05)**   | **0.10(0.07)**   |
|                        | coefficient     |                  |                  |                  |
| Topological profiles match disease progression better than any single descriptor |
|-------------------------------|-------------------|-------------------|-------------------|
| The third row of Figure 2 shows predicted atrophy progression using the best-matching single topological descriptor. Overall, the combinations of descriptors (Figure 2, second row) match the data (Figure 2, first row) more closely than the single best-fitting descriptor (Figure 2, third row). For example, according to the topological profile prediction, the parietal lobe is involved in the early stages of AD, in agreement with the data, while it appears to be involved at a later stage, according to the single best-fitting descriptor (network proximity, underlined in Table 1), which also underestimates the involvement of the subcortical areas. Similarly in PPMS, the topological profile prediction reproduces the subcortical involvement better than the best-fitting single descriptor (segregation – inverse clustering). In HA the topological profile prediction shows involvement of the temporal and frontal lobes, as does the atrophy pattern, while the strongest single descriptor (cortical proximity) underestimates subcortical involvement and overestimates parietal involvement. Further, we note that topological profiles explain variance in the data better than the best-fitting single descriptors. Indeed, in the AD cohort, the topological profile explains 82% of the variance, with the constant term explaining just 6%, in contrast to 51% explained by network proximity. In the PPMS cohort, the topological profile explains 83%, the constant term 7%, and inverse clustering 25%. Similarly, in | [0.21(0.12)] | [0.14(0.06)] | [0.08(0.06)] |
| {0.01(3.32)} | {0.01(4.52)} | {0.01(2.97)} |
| Segregation | Inverse degree | 0.05(0.06) | **0.17(0.12)** | 0.01(0.01) |
| | | [0.05(0.05)] | [0.016(0.11)] | [0.01(0.01)] |
| | | {0.05(0.26)} | {0.01(3.43)} | {0.01(5.31)} |
| | Inverse clustering | 0.01(0.03) | **0.32(0.22)** | 0.01(0.01) |
| | | [0.05(0.04)] | [0.36(0.24)] | [0.01(0.02)] |
| | | {0.01(1.90)} | {0.01(2.82)} | {0.01(2.01)} |
| Network proximity | Shortest path | **0.65(0.39)** | 0.06(0.06) | 0.01(0.04) |
| | | [0.54(0.35)] | [0.07(0.06)] | [0.01(0.02)] |
| | | {0.01(3.62)} | {0.01(2.48)} | {0.01(1.48)} |
| Cortical proximity | Spatial distance | 0.06(0.08) | **0.22(0.14)** | **0.64(0.38)** |
| | | [0.10(0.06)] | [0.21(0.18)] | [0.54(0.32)] |
| | | {0.01(2.14)} | {0.01(3.22)} | {0.01(3.46)} |
| Constant progression | Constant term | 0.07(0.02) | 0.08(0.04) | **0.19(0.09)** |
| | | [0.12(0.03)] | [0.10(0.05)] | [0.18(0.12)] |
| | | {0.01(4.01)} | {0.01(3.56)} | {0.01(3.58)} |

Table 1. Weights of the topological profiles of the three cohorts. The table reports the weights for each network metric, grouped per topological descriptor. Credible intervals for the weights are given in parentheses. Bootstrapping variation is shown in square brackets. Bonferroni-corrected $p$-values and effect size for the permutation testing of the null hypothesis are shown in braces. In bold the weights that have been used to compute the topological profiles.
the HA cohort the topological profile explains 88%, the constant term 16%, and cortical proximity 64%. To quantify and compare how well the topological model predictions match the data, we also calculate the Akaike Information Criterion (AIC), which penalizes model complexity. The topological profiles always provide lower AIC scores than the best-fitting single descriptor model (see Supplementary Information, Table S4), thus explaining the data better without overfitting. For full details please refer to the Methods section.

**Individual profiles group around cohort profiles and separate in topological space**

Figure 3 plots the cohort-level topological profiles and the individual topological profiles on different topological spaces, for different groups of individuals. Personalized topological profiles come from fitting the best combination of network metrics to individual progression rates. The GP Progression Model provides individual time parameters positioning the individuals along the estimated progression and thus enables the estimation of individual progression rates with respect to the global disease progression (see Methods: Personalized topological profiles).

We consider a 5D topological space spanned by each of the five descriptors we consider (centrality, segregation, network proximity, cortical proximity and constant progression). The position of a particular topological profile within this space has coordinates that are the sums of the weights of the network metrics corresponding to each descriptor. The ternary plots in Figures 3(a), 3(c) and 3(e) show the distribution of individual profiles colored according to the cohort to which the individual belongs (red for AD, green for PPMS, blue for HA). The position corresponds to the relative Euclidean distances of the individual profile from each cohort-level profile in the 5D topological space. Thus the corners correspond to exact matches with one of the cohort profiles, while points at the center of the triangle are equidistant to all three cohort-level profiles. Individual profiles that are closer to the profile of another cohort are highlighted as black diamonds with border color reflecting the true cohort; we refer to these individuals as “outliers”. Figures 3(b), 3(d) and 3(f) show projections of the 5D topological space to 2D using t-Distributed Stochastic Neighbor Embedding (tSNE: ), which visualizes high dimensional points in a low dimensional space in a way that retains pairwise similarity with high probability, i.e. most points that are close/distant in the 5D space are close/distant in the 2D visualization. Thus, while the global shapes of the distributions of points have limited interpretation, disconnected groups of points in the tSNE plot reflect separation in the native space. Each tSNE plot shows the cohort-level topological profile from each full cohort (big stars), the variation of the cohort-level topological profile under bootstrap resampling (small stars), and each individual topological profile (dots).

Figures 3(a) and 3(b) show the plots for topological profiles computed using the GP Progression Model. They include points for only disease-diagnosed individuals in the disease cohorts (AD or MCI in the AD cohort; PPMS-diagnosed in the PPMS cohort; excluding controls in both), but all individuals from the HA cohort. Figures 3(c) and 3(d) show only the controls from the AD and PPMS cohort, as well as all the HA cohort, also using the topological profiles obtained from the full temporal trajectories. Figures 3(e) and 3(f) show the same individuals as Figures 3(a) and 3(b) but using topological profiles derived from only late-stage patient data, mimicking current standard models, which ignore disease progression; see Supplementary Information: Late-stage atrophy modeling. Supplementary Table S3 shows results using random networks to show that the separation is genuinely driven by the underlying structural connectivity.

Figures 3(a) and 3(b) show that i) the cohort topological profiles are consistent under bootstrap cross-validation, ii) the three cohorts separate well in topological space, and iii) individual profiles group around cohort profiles.

Figures 3(c) and 3(d) and supplementary Table S4 confirm that the separation observed in Figures 3(a) and 3(b) is due to disease-related information, and not to differences in MRI centres, MRI scanner or
acquisition protocol, as the three groups of healthy controls show weaker separation than the distinct disease groups.

Matching progression pattern defines profile better than late-stage atrophy alone

Figures 3(e) and 3(f) show that when using only late-stage atrophy to identify topological profiles: i) the cohorts separate less strongly than using the DPM topological profiles (see Table S2); and ii) the rate of assignment of AD and PPMS individual profiles to the corresponding cohort profile is lower than using GP Progression Model-driven topological profiles and with a greater fraction of outliers (see Table 3). Both observations suggest that late-stage information alone is not sufficient to provide distinct topological profiles.

Table 2 shows the weights for the topological profiles when using late-stage data. We observe that, with respect to the topological profiles estimated using the GP Progression Model (Table 1), the centrality components in AD and HA are decreased (from 40% to 25% and from 25% to 15%, respectively), including an increase in uncertainty for all the estimates. In HA we also note an increase in the constant propagation term (from 15% to 30%), which may be explained as a compensatory effect for the decrease of the other weights, as the constant propagation term is analogous to the intercept in linear regression – for details see Methods section.

Table 3 is a confusion matrix of classification rates for subject assignments: without parentheses are the rates for assignment using the GP Progression Model (using patient data only — no controls); numbers in parentheses are the rates when using late-stage information (patients only). These confirm that superior classification rates come from using GP Progression Model driven profiles.

| Topological descriptors | Network metrics | AD       | PPMS     | HA       |
|-------------------------|-----------------|----------|----------|----------|
|                         | Betweenness centrality | 0.15(0.11) | 0.11(0.09) | 0.05(0.02) |
| Centrality              | Closeness centrality | 0.01(0.03) | 0.06(0.07) | 0.01(0.02) |
|                         | Weighted degree   | 0.03(0.04) | 0.11(0.08) | 0.09(0.05) |
|                         | Clustering coefficient | 0.07(0.10) | 0.11(0.15) | 0.03(0.02) |
| Segregation             | Inverse degree    | 0.06(0.06) | 0.12(0.09) | 0.06(0.06) |
|                         | Inverse clustering | 0.09(0.08) | 0.25(0.22) | 0.01(0.04) |
| Network proximity       | Shortest path     | 0.30(0.22) | 0.05(0.05) | 0.02(0.04) |
| Cortical proximity      | Spatial distance  | 0.12(0.09) | 0.15(0.09) | 0.15(0.09) |
| Constant progression    | Constant term     | 0.04(0.05) | 0.10(0.07) | 0.33(0.21) |

Table 2. Weights of the topological profiles of the three cohorts when using only late-stage atrophy data are more uncertain and overlap more. The table reports the weights for each network metric, grouped per topological descriptor.
Table 3. Confusion matrix of classification rates for individuals’ assignment to each cohort by matching individual topological profiles to cohort-level topological profiles. Without parentheses: results using full GP Progression Model-driven topological profiles. Within parentheses: results using topological profiles estimated from only late-stage data. Higher numbers are better in diagonal entries (correct assignment), and lower numbers are better in off-diagonal entries (incorrect assignment).

|       | AD     | PPMS   | HA     |
|-------|--------|--------|--------|
| AD    | 84% (57%) | 4% (9%) | 12% (34%) |
| PPMS  | 2% (16%) | 68% (45%) | 29% (38%) |
| HA    | 7% (14%) | 2% (12%) | 91% (74%) |

Alzheimer’s disease individuals that exhibit individual profiles closer to the healthy aging topological profile perform better in cognitive tests

Figure 3(a) identifies 159 AD-cohort outliers (130 MCI, 29 probable AD) closer to the HA cohort-level profile, out of 1312 MCI+AD patients. We analyzed clinical and demographic information for the AD outliers with respect to the rest of the AD cohort (see Supplementary Table S5a). An ANOVA 1-way test between the two groups was performed and differences were found in MMSE, where the outliers exhibit a higher MMSE score (27.6±1.2 vs 26.1±1.7). Slight differences were found in the Clinical Dementia Rating Scale, with outliers scoring lower (1.8±0.7 vs 2.1±0.3). Both results show that the outliers have reduced cognitive deficits compared to inliers, on average. Further information on the ANOVA results is provided in the Supplementary Table S5a. The same analysis performed on the outliers defined by late-stage-driven topological profiles did not report any significant differences between the groups, suggesting that outliers defined using the GP Progression Model-driven topological profiles depart more genuinely from the disease cohort and phenotype of typical amnestic AD; see the Supplementary Table S6a).

Healthy ageing individuals that exhibit individual profiles closer to the AD topological profile show signs of prodromal dementia

Figure 3(a) identifies 358 of 5463 HA individuals closer to the AD cohort profile. An ANOVA 1-way test between these outliers and the rest of the HA group found differences in MMSE, where the outliers exhibit a lower MMSE score (27.7±0.8 vs 28.0±0.6); and a difference in age, with the outliers being older (70.5±10.1 vs 64.5±9.8). No significant difference in gender or APOE status (see Supplementary Table S5c) was found. In order to investigate whether the increased cognitive deficit in the HA outliers indicates an actual prodromal phase of dementia, or is just an age-related effect, we analyzed the incidence, in the outlier group, of the individuals that were healthy at baseline, but developed dementia after 2–4 years (see Methods: Participants - HA). Of those 148 individuals, 105 reside in the HA-AD outlier group, which thus contains 70% of the individuals with prodromal dementia; in contrast the non-outlier group consists of only 5% prodromal dementia cases. The same analysis performed on the outliers defined by late-stage-driven topological profiles reports only a group difference in age with older outliers (65.2±10.9 vs 60.0±7.0). No other significant differences were found (see Supplementary Table S6c)).

Primary progressive multiple sclerosis outliers show no significant differences with the rest of the group

We also analyzed PPMS with respect to both AD and HA: the analysis returned 53 AD individuals closer to PPMS; 1 PPMS individual closer to AD; 13 PPMS individuals closer to HA and 100 HA individuals closer
to PPMS. No significant differences were found between any subgroups, for any comparisons: we analyzed age, gender, expanded disability status scale (EDSS) and disease duration when looking at PPMS outliers in AD or HA; age, gender, APOE4 and MMSE score when looking at HA outliers in PPMS; age, gender, years of education, MMSE, ADAS-Cog, APOE4, CDRSB, AV45, when looking at AD outliers in PPMS. Details can be found in Supplementary Table S5b). The same analysis on late-stage-driven topological profiles also returned no significant differences (Supplementary Table S6b).

Features derived from the topological profile correlate with clinical features

Finally and more generally, we analyzed features extracted from topological profiles, in particular, the distance of individual profiles from the cohort profile, and show that they correlate negatively with individual cohort-specific clinical features. Specifically, MMSE for the AD cohort (R=0.11, p<0.01); EDSS for PPMS (R=0.68, p=0.07); and age for HA (R=0.32, p<0.01).

Discussion

We have introduced a novel method to estimate cohort and individual-level topological profiles of neurodegeneration using computational disease progression models in combination with imaging data sets. The profiles give new insight into the relationship between brain connectivity and the progression pattern of neurodegeneration. We demonstrate the ideas using three cohorts representing different neurological conditions: AD, PPMS, HA. We showed that combinations of topological descriptors explain observations better than any individual descriptor and that the combinations representative of each condition, despite some commonality as suggested in \(6^{–8}\), differ substantially both at the cohort and individual level. We emphasize that these differences go beyond simple observations that the pattern of atrophy accumulation is different in these three cohorts, which could arise simply from equivalent spreading mechanisms from different epicentres; our results go further by strongly suggesting distinct modes of dependence of the pathological spread on the underlying connectivity. We also show that using the full disease time-course, estimated via a GP Progression Model \(^{59}\) produces better-defined topological profiles than using only late-stage atrophy. This highlights a key weakness in previous state-of-the-art results using late-stage atrophy alone, which can be ameliorated using recent advances in disease progression modeling. Further, we retrieved significant correlation between features of the topological profiles and individual clinical or demographic features, suggesting potential clinical utility of the topological profile. Finally, we showed that the outliers in the AD-HA comparison display characteristics that align them with the other cohort, providing evidence that our topological profiles reflect underlying disease processes, and indicating potential use of the topological profile to highlight disease risk.

As discussed in the introduction, one key implication here (as in previous literature\(^{3,20,68,69}\)) is that graph-theoretic measures may associate with different mechanisms of disease spread in neural networks. Specifically, centrality represents nodal stress (vulnerability of brain hubs), segregation trophic failure (vulnerability of isolated regions), network proximity transneuronal spread of a prion-like agent along network connections, and cortical proximity unguided diffusive propagation \(^5\). While we draw on these potential links in the following sections, it is important to note that the links between the mathematical abstraction and the biological mechanisms are simplistic, and graph-theory metrics do not fully capture or explain the range of potential spreading mechanisms.

The centrality component in AD topological profiles becomes apparent when taking into account the temporal evolution of atrophy

The AD topological profile identified here supports recent results that identify neuronal distance from epicentre (or trans-neuronal spread) as the principal topological descriptor of neurodegeneration in AD \(^{7,14,17,20,23}\). However, both cohort- and individual-level profiles also show a substantial additional component
of neurodegeneration due to centrality metrics — vulnerability of brain network hubs. This component has
not clearly been highlighted previously, because a) it is less apparent in a topological profile estimated from
only late-stage atrophy data, and b) it is suppressed when using the previous single-descriptor approach. We
found contrasting importance of centrality metrics between full-disease-course and end-stage-only profiles.
This might arise from metabolic demand increasing primarily during early degeneration so that hub-
vulnerability drives early phases of the progression pattern\(^70\), while neuronal distance from the epicentre
dominate later stages. This finding might also explain some aspects of the biological heterogeneity of AD,
which produces high variance in atrophy patterns across individuals. In particular, the early centrality
component of the topological profile suggests that pathology may start around local hubs, which is consistent
with the focal presentation of atrophy in atypical Alzheimer’s diseases such as posterior cortical atrophy or
progressive aphasia\(^71,73\). Once established, each disease then spreads through the same broad networks,
leading to syndromic convergence in advanced stages of AD.

Similarly, in HA and PPMS, hub vulnerability appears as a significant component only when
considering the full trajectory rather than only late-stage atrophy. This may suggest that activity-dependent
mechanisms, such as oxidative stress\(^74\), microglial activation\(^75,76\), neurovascular dysfunction\(^37\) or “virtual
hypoxia”\(^38\) influence early brain loss across a range of neurodegenerative conditions.

More generally, the differences we observe between full-disease-course and end-stage-only profiles
emphasize the importance of considering the full temporal trajectory in assessing the likelihood of potential
propagation mechanisms.

**Degree of alignment of individual profiles with cohort profiles informs on patient status**

Our results indicate that departures of individual profiles from that of the corresponding cohort may signal
abnormality with respect to their cohort, while alignment with another cohort-level profile may correspond
to clinical features of that other cohort. For example, the AD individuals with a topological profile close to
the HA cohort-level profile show on average lower cognitive deficit than the full AD cohort. On the other
hand, HA individuals with topological profiles that tend towards the AD cohort-level profile exhibit reduced
MMSE score, suggesting cognitive impairment; also, a substantial proportion (70%) of the HA prodromal-
dementia individuals (who developed dementia in 2-4 year follow-up) lie in the outlier group, suggesting
evidence of neurodegenerative mechanisms consistent with AD. We emphasize again that we identify these
outliers using the topological descriptors that explain the atrophy pattern, and not on the atrophy pattern
itself, with the intention of highlighting individuals showing abnormal mechanisms of pathology propagation
with respect to their cohort.

We found no significant difference in clinical features between PPMS outliers, tending towards
either AD or HA, compared with the rest of the PPMS group: topological profiles seem to identify
differences among PPMS subjects that are invisible to the available demographic and clinical features (age,
gender, EDSS and diseases duration). However, sample sizes are low (N=1 PPMS-AD outlier; N=13 PPMS-
HA outliers). Further investigation on a larger cohort of progressive MS subjects, as well as with a broader
class of clinical features, may clarify whether the variation we observe in topological profiles arise from
spurious effects, or are genuinely informative on the clinical status of the outliers beyond what the available
clinical features can identify. For example, future studies of older and longer-term PPMS patients could test
the hypothesis that individual topological profiles may inform on co-morbidities, e.g. for an MS-patient who
shows signs of dementia, their profile position with respect to the AD cohort profile may inform on whether
the dementia comes from MS-related neurodegeneration or AD comorbidity. However, we cannot test this
hypothesis here, as our PPMS cohort are “early” PPMS (they were recruited to this study within 5 years of
diagnosis), and no subject reported dementia as would be expected\(^77\) although a detailed cognitive
assessment was not performed, this agrees.
Neurodegeneration in AD and PPMS is not simply an acceleration of the aging process

Overall, our results suggest that neurodegeneration in AD and PPMS is not simply an acceleration of the aging process, but that the two diseases have distinct topological profiles of neurodegeneration, which differ from those of HA. The cohort and individual-level topological profiles we identify for AD and PPMS are clearly distinct from one another. The AD topological profile is a combination of centrality and network proximity, while the PPMS profile has diverse contributing mechanisms including centrality, segregation and cortical proximity, suggesting that many different processes are at play. Also, both separate from the HA topological profile, which has a strong component of both the cortical proximity and the constant propagation mechanisms, suggesting relatively uniform loss across the brain that is not linked to any particular brain-connectivity feature.

Further, we note that we have regressed out the effects of ageing in the AD and PPMS cohort in order to minimize the effects of the ageing process on disease models. Nevertheless, some components of ageing might still appear in the model for both, which is one key reason for including the HA model for comparison.

These observations underline the idea that mechanisms, in addition to actual atrophy patterns, are distinct in these diseases from those underpinning normal aging.

Future Work

Here we provide a proof of concept of the idea of topological profiling through image-analysis and temporal disease progression modeling, and a demonstration of its potential utility. Multiple opportunities arise for future refinements of the methods, as well as extensions and applications of the key ideas.

Methodologically, several steps in the processing pipeline can limit the accuracy of the network metrics in representing neurodegenerative mechanisms. For example, structural connectome estimation using tractography is prone to false positive and negative connections that can influence subsequent predictions. Nevertheless, tractography does reliably recover at least many of the major known anatomical connections sufficiently well to highlight broad differences in atrophy patterns associated to the different mechanisms. Anatomically-constrained probabilistic tractography coupled with SIFT-ing should provide more accurate predictions than deterministic tractography, but future work might consider better ways to mitigate errors e.g. by quantifying uncertainty in tractography output.

Here we take an average connectome over multiple young and healthy individuals as the substrate for pathology propagation. By doing so, we do not account for any influence of the pathology on the connectome itself, as some other models attempt to. However, we believe this to be a reasonable anatomical reference of a “pure” underlying substrate of propagation to investigate grey matter loss and decide to avoid deliberately the additional complexity of white matter disruptions, which are not well understood and would complicate the model substantially. This decision is also supported by recent results in AD showing that the choice of connectome (young and healthy or subject-specific) does not significantly impact the predictive ability of a model for trans-synaptic transmission of toxic proteins. However, future models could approximate reductions of connectivity arising from white matter damage commensurate with grey-matter atrophy, but this requires a clear picture of the interaction of such processes, e.g. from disease progression models of regional grey-matter atrophy and white-matter integrity. We believe using the healthy connectome as the substrate for propagation is a sensible first-order approximation.

One major advantage of our approach for evaluating topological descriptors is that it considers group-level longitudinal information on atrophy rather than considering only late-stage atrophy patterns. Our method uses each network metric to obtain a scalar value at each graph node that provides a template of the rate of pathology accrual at that node compared to others, and then compares to rate-of-change of atrophy
evolution as estimated by the GP Progression Model. Future work may extend the idea by accounting for variable change over time, which will need to compare the full temporal evolution of atrophy to the full temporal evolution induced by each topological descriptor. One additional caveat of using the GP Progression Model is that the model captures the trajectory of change in biomarker values rather than underlying pathology; differences among biomarker accuracy (e.g. higher noise in volume measurements of smaller regions) can make the order of biomarker change depart from the order of underlying pathological change, although we expect such effects tend to be small with our processing pipeline – see \(^{44,55}\) for further discussion.

Our approach does not account for heterogeneity within cohorts. In addition to variation in the structural connectome discussed above, the epicentre for disease propagation may vary among patients in a particular disease class and a number of recent works explore methods for per-subject epicentre selection \(^{54,87}\). Here we estimated a single cohort-level epicentre as the most atrophied region for each data set, and each reflected common knowledge of brain tissue loss in AD, PPMS and HA. Our epicentres broadly agree with \(^5\), who found that individually-selected epicentres in a number of neurodegenerative diseases are always seated in or near the most atrophied region. Recent data-driven approaches \(^{88}\) further reveal within-cohort heterogeneity of temporal trajectories, which our current approach does not accommodate, but such advances provide great potential for future work to focus topological profiling on more homogeneous subgroups and to explore the variability of topological profiles among different disease subgroups. Indeed, we believe application to broader cohorts, as for instance generic progressive MS including both primary and secondary (which would be more numerous) would require better identification of distinct subgroups than the traditional MS classifications (clinically isolated syndrome, relapsing remitting MS, and primary and secondary progressive MS), and methods such as SuStain \(^{89}\) would offer a way to identify within-cohort subtypes for better topological profile estimation. As concerns applications to prodromal phases of diseases – an example being radiologically isolated syndrome for MS – future work will be devoted to build models of the whole neurodegeneration process within specific diseases and their subtypes defined by distinct trajectories of pathological change, which might be used for instance for discriminating subjects that convert to degeneration from others that do not, thereby informing treatment and care choices.

The ideas we propose here extend to a much larger range of diseases and disease subtypes and offer a new and unique tool to exploit the full power of large imaging data sets in exposing mechanisms of disease aetiology and propagation. With further development and experiments, topological profiles may also provide useful information in the clinic. For example, AD-like HA outliers may be considered at risk for AD and decisions could be informed by this, including patient management (more frequent visits to monitor the patient) and recruitment into clinical studies/trials.

The underlying models of propagation remain simplistic and a great many alternative descriptors and mechanisms could easily be included, such as propagation via functional networks \(^{5,17}\), or different kinds of tractography to represent intra and extra-axonal propagation \(^{25,26,51}\). Simplicity of the models and limitations of processing steps is important to bear in mind while interpreting results, and validation, e.g. against histological analysis in animal models \(^{89}\), is an important future target to establish the extent to which mechanistic information inferred in this way matches low-level observations of molecular propagation.

**Materials and Methods**

Our pipeline for creating topological profiles in a neurodegenerative condition is the following.

1) Generate an average structural connectome from young healthy data.
2) Pre-process the cohort data to adjust for nuisance variables and generate the longitudinal data for regional atrophy and topology, relative to healthy controls.
3) Build the DPM and topological profile.
We now describe the data and these steps in detail.

Data Description

Participants – AD

The data used in the preparation of this article were obtained from the Alzheimer’s disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org. We collected longitudinal measurements for all the available ADNI 1/GO/2 individuals with at least one “quality control” flagged 3D-T1 MRI scan. Repeated T1-weighted structural MRI images were acquired at 3T machines across multiple centers according to a harmonized protocol. Longitudinal FreeSurfer was used to align images from multiple time-points according to subject specific median templates, in order to avoid temporal bias. Subject with no available diagnosis were discarded, thus leaving us 1713 subjects, with “HC”, “MCI” or “AD” diagnosis. We collected longitudinal FreeSurfer 5.1 volume data on all the GM and subcortical regions from ADNIMERGE.csv. For information on scanning protocols and segmentation algorithms see www.adni-info.org. The subjects included in our analysis are 1713; the mean (std) age of the cohort is 73.9(7.2) years. The overall scans are 6670; the average (std) time between scans is 2.4(1.8) years. More information are in the Supplementary Information (Figure 3-figure supplement 1 and Table S7).

Participants – PPMS

This was a retrospective study of 64 participants, studied at the UCL Queen Square Institute of Neurology in London. The participants include 44 individuals with primary progressive multiple sclerosis (PPMS) and 20 healthy controls (HC). We collected longitudinal FreeSurfer 5.1 volumetric data on all the GM and subcortical regions. For information on scanning protocol, segmentation algorithms, please refer to 90. The longitudinal data set used for this study consisted in 64 individuals; the mean (std) age of the cohort is 41.6(10.2) years. The overall scans are 244 scans; the average (std) time between scans is 4.0(1.5) years. More information are in the Supplementary Information (Figure 3-figure supplement 1 and Table S8).

Participants – HA

The data used in the preparation of this article were obtained from The Rotterdam Scan Study.91,92 The Rotterdam Study is a prospective cohort study ongoing since 1990 in the city of Rotterdam in The Netherlands. The study targets cardiovascular, endocrine, hepatic, neurological, ophthalmic, psychiatric, dermatological, otolaryngology, locomotor, and respiratory diseases. Initially, in 1995 and 1999, random subsamples of participants from the Rotterdam Study underwent neuroimaging, whereas from 2005 onwards brain MRI has been implemented into the core protocol of the Rotterdam Study. We excluded individuals with prevalent dementia at study entry, and individuals presenting cortical infarcts on the MRI. The study contains 148 individuals that would develop dementia in the 2-4 years follow-up(s). We leave them in the study as we are interested to model their individual topological profiles with respect to the topological profile of the cohort-level. We collected longitudinal FreeSurfer 5.1 volumetric data on all the GM and subcortical regions. For information on the scanning protocols and segmentation algorithms see 92. The longitudinal data set used for this study consisted in 5463 individuals; the mean (std) age of the cohort is 64.8(10.8) years. The overall scans are 11627 scans; the average (std) time between scans is 5.3(1.1) years. More information are in the Supplementary Information (Figure 3-figure supplement 1 and Table S9).
Data pre-processing to obtain model inputs

On each data set of volumetric GM and subcortical regions separately, we performed three steps:

1) Adjustment for nuisance variables: we constructed a regression model for each region separately with volumes as dependent variable and total intracranial volume, gender, and age (age not included for the HA cohort) as independent variables.

2) Variable selection: from the original FreeSurfer Regions of Interest (ROIs) we discarded white matter, brain stem, ventricular and cerebellar regions, leaving 82 ROIs. Then, we averaged the (adjusted) volumes of each region from both hemispheres, obtaining 41 bilateral regions.

3) Z-scores computation: we computed z-scores against a control population. For the AD data set, the control population was formed by the “HC”-diagnosed individuals; for the HA cohort it was formed by the “Young” individuals (those whose age is more than one standard deviation less than the mean age); for the PPMS data set it was formed by the “HC”-diagnosed individuals.

HCP participants

Data used in the preparation of this work were obtained from the MGH-USC Human Connectome Project (HCP) database (ida.loni.usc.edu/login.jsp). The HCP project (Principal Investigators: Bruce Rosen, M.D., Ph.D., Martinos Center at Massachusetts General Hospital; Arthur W. Toga, Ph.D., University of California, Los Angeles, Van J. Weeden, MD, Martinos Center at Massachusetts General Hospital) is supported by the National Institute of Dental and Cranio-facial Research (NIDCR), the National Institute of Mental Health (NIMH) and the National Institute of Neurological Disorders and Stroke (NINDS). Collectively, the HCP is the result of efforts of co-investigators from the University of California, Los Angeles, Martinos Center for Biomedical Imaging at Massachusetts General Hospital (MGH), Washington University, and the University of Minnesota. The data set consisted of 24 unique subjects: we collected raw high-resolution 3D T1–weighted and DTI of 24 age and gender-matched subjects (age: 26, 50% female).

Connectome generation

Structural connectomes were generated using tools provided in the MRtrix3 software package (http://mrtrix.org). The pipeline included: DWI denoising; pre-processing and bias–field correction; inter–modal registration; T1 tissue segmentation; spherical deconvolution; probabilistic tractography utilizing anatomically-constrained tractography; dynamic seeding and SIFT; T1 parcellation; robust structural connectome construction. Our anatomical connectome for each participant is a weighted adjacency matrix that includes only inter-node connections across the 82 ROIs consisting of cortical and subcortical gray-matter regions, excluding the cerebellum and brain stem. The average structural adjacency matrix was computed by taking the mean over the subject-wise matrices. Weights, or connection strengths, were normalized to [0, 1]. The inter-hemisphere average is performed after the network metrics computation step (see below).

Mathematical modeling

Network metrics

We selected five different descriptors of graph topology (centrality, segregation, network proximity cortical proximity and constant progression) and 9 metrics across these categories:

1. Centrality measures:

   a) Weighted degree: the sum of the weights of the connections incident the node of interest.
b) Betweenness centrality: the number of shortest paths between any two nodes that pass through the specific node.

c) Closeness centrality: the inverse of the path length between the node of interest and all the other nodes.

d) Clustering coefficient: the proportion of triangular sub-networks formed by the node and its neighboring nodes.

2. Segregation measures:

a) Inverse of weighted degree.

b) Inverse of clustering coefficient.

3. Network proximity:

a) Shortest path to epicenter.

4. Cortical proximity:

a) Spatial distance from epicenter.

5. Constant progression:

a) Constant term quantifying to what extent the rate of atrophy remains constant throughout the progression.

Each network metric produces a scalar value for each node that indicates its vulnerability to pathology relative to all other nodes. Each thus leads to a temporal progression pattern in which the accumulation of pathology, e.g. the atrophy or rate of change of volume, is proportional to that vulnerability (see Figure 4). In order to identify the epicenters needed to compute 3a) and 4a), we selected, for each subject, the region with the highest z-scores of GM volumes (i.e. the “most atrophied” region) in the last follow-up MRI scan; the cohort-level epicenters are then, for each cohort, the regions that most frequently appear as individual-level epicenters. They are the hippocampus (AD), the caudate (PPMS) and the insula (HA). These epicenters are supported by literature in AD, PPMS and HA.

We computed the first 7 network metrics using the Brain Connectivity Toolbox in MATLAB, after appropriate normalization (weight_conversion function). To compute the cortical proximity metric, which does depend on the structural connections, we select the segmented structural MRIs in the HCP cohort, and measure the (average) pairwise cortical distances between all the macro-regions’ barycenters (using the mris_pmake command in FreeSurfer). The constant progression metric, which also does not depend on the structural connections, simply assigns a constant value to each region.

We note that the network metrics computing similar attributes of a graph such as centrality and efficiency, or segregation and inefficiency, are not independent and in fact highly correlated producing strongly overlapping progression patterns. However, we include multiple network metrics for certain topological descriptors to capture variability in the corresponding progression pattern that fits within the definition of the descriptor.

Network metrics were computed on the average structural adjacency matrix formed of 82 ROIs, and then averaged across hemisphere, obtaining an array of 9 metrics for each of the 41 symmetrized regions \( v, g(v) = (g_1(v), ..., g_9(v))^T \), where we set \( g_9(v) = 1 \) to describe the constant progression term.
GP progression model

Continuous Disease Progression Models (DPM) aim to estimate the long-term temporal pattern of disease progression directly from short-term longitudinal data sets and to stage patients based on individual observations. The problem is challenging, due to the lack of a well-defined temporal reference in longitudinal data sets: time of onset is usually unknown and rate of pathology accrual is highly variable in most neurodegenerative conditions. DPM techniques typically tackle this problem by assuming that each visit at which measurements are taken occurs at an unknown “disease time”, the particular value of which for each individual subject is a hidden variable estimated while fitting the set of trajectories. The disease-time axis parametrizes the natural evolution of the pathology common to all individuals.

The models usually assume mixed effects, in which the long-term trajectories are the fixed effect (represented by parametric or non-parametric functions), and the individual variations from the group-level trajectories are the random effects. Further, individual time reparametrization parameters (also known as time warp parameters$^{55}$ or disease scores$^{41,46}$) are estimated by quantifying the individual positions with respect to the estimated time-frame. Here, we use the GP Progression Model of$^{59}$, which is a non-parametric Bayesian mixed effect model, to estimate long-term trajectories of regional GM volumes on each data set (AD, PPMS and HA) separately. Figure 5 shows a pictorial example of the GP Progression Model combined with the topological profiles procedure.

Formally, if we represent by $(X^j(t_1), ..., X^j(t_K^j))$ the longitudinal measurements of the regional GM volumes associated with each individual $j$ at their $(t_1, ..., t_K^j)$ absolute time points (i.e. dates of subject visit or subject’s age-at-visit), and consider all measurements obtained at a particular visit of one individual to occur at a particular disease time $t_K^j$, where the mapping from absolute time $t$ to disease time $\tau$ is via a subject-specific time reparametrization function, which here is just a simple shift: $t_K^j = \tau_K + d^j$, then the observations for subject $j$ at a single time point $t$ (indices omitted but implied) can be modelled as a random sample from the GP Progression Model:

$$X^j(t) = \left( X_1^j(t), ..., X_V^j(t) \right)^T = f(t) + n^j(t) + E. \#(1)$$

Here $f(t) = (f_1(t), ..., f_V(t))^T$ is the fixed effect function modeling the longitudinal evolution of the $V$ GM volumes, and is modeled as a Gaussian Process; $n^j(t) = (n_1^j(t), ..., n_V^j(t))^T$ are the individual random effects; and $E = (E_1, ..., E_V)^T$ is the observational noise. The model is described in detail in$^{59}$ together with the optimization scheme to recover the probabilistic estimates of the parameters for the fixed effect, the random effect, and the individual time reparametrization parameters. Identifiability of the model is ensured by enforcing monotonicity on the population-level biomarker trajectories $f_v$.

Fitting the model also provides an estimate of the highest rate of change of each biomarker during the disease progression. That is the maximum of the derivatives of the estimated trajectories $f_v$ along the disease time $\tau$:

$$Y_v = \max_{\tau} f_v'(\tau), \#(2)$$

for each biomarker $v \in \{1, ..., V\}$. Existence and boundness of $Y_v$ is guaranteed by the finiteness of the time-shifts, which is enforced by the smoothness of the Gaussian Process. Indeed, for each subject $j$, $d^j$ defines the optimal shift of the data point on the temporal time axis. The estimates of these positions must be compatible with the Gaussian Process describing the temporal trajectory (Figure 2-figure supplement 2, 3 and 4). Gaussian processes are completely identified by the kernel function (in our case a radial basis function - RBF), which prescribes the shape and smoothness of the interpolating curve via its length-scale $l$.
and variance $\sigma$. For this reason, the relative positions (i.e. the time-shifts) of each individual are naturally bounded by the length-scale of the Gaussian process, so the only compatible solutions are those with the time-shifts softly constrained to a finite range determined by $2l$. The monotonicity constraint on $f_v$ guarantees uniqueness of $Y_v$ for each $v$. The model returns uncertainty on $f$, which can be projected to $f'$ and thus $Y$.

**Topological profile estimate**

We estimate the topological profile $\beta$ by identifying the unique combination of topological descriptors that best matches the GP Progression. This means estimating the weights $\beta$ from the linear model

$$Y = \Sigma \beta + \epsilon, (#3)$$

where $Y = (Y_1, ..., Y_V)^T$; $\Sigma$ collects the values of the metrics $g(v)$ for every region $v$, i.e. $\Sigma = (g(1), ..., g(V))^T$; and $\epsilon$ is the noise. As a first step we normalize both $Y$ and $\Sigma$ to the range $[0, 1]$. We also enforce non-negativity of $\beta$, so that $\beta_v$ can be precisely interpreted as the weight with which descriptor $v$ contributes to the overall observed pattern of neurodegeneration $Y$. Our problem is then the one of estimating $\beta$ such that:

$$\beta = \arg\max_{\beta \geq 0} L(Y, \Sigma \beta), (#4)$$

where $L$ is the likelihood of the model, which assumes Gaussian noise, so that (4) becomes equivalent to constrained least-squares minimization. We solve the problem via Expectation-Maximization, which, in the positively-constrained least-squared case, has a simple closed form, and becomes an iterative-multiplicative algorithm also known as ISRA\textsuperscript{110,111}.

**Personalized topological profiles**

For each subject, the GP Progression Model estimates a set of time-points $t^j_k$ describing the subject’s measurements in the new disease time $\tau$ as $t^j_k = \tau^j_k + d^j$, where $d^j$ is the subject-specific shift. If we define the subject temporal position in the disease time as the shifted average age $\tilde{\tau}^j$, where $\tilde{\tau}^j = \tau^j + d^j$ and $\tilde{\tau}^j$ is the average age of subject $j$ across age-at-visits, then we can compute the individual speed of progression along the disease time as

$$Y_v^j = f_v^j(\tilde{\tau}), \quad (#5)$$

for each biomarker $v \in \{1, ..., V\}$. These values encode information on the individual rates of change along the disease progression. We can then estimate the individual topological profile $\beta^j$ by identifying the unique combination of topological descriptors that best matches the individual progression as

$$\beta^j = \arg\max_{\beta \geq 0} L(y^j, \Sigma \beta^j), (#6)$$

where $y^j = (Y_1^j, ..., Y_V^j)^T$. As in the cohort-level framework, the problem can be solved, for each subject $j$, via ISRA.

**Model selection**

In order to quantify the performances of the topological profiles progression against the single-best fitting descriptor in reproducing the observed progression from the GP Progression Model, we compute Aikake Information Criterion (AIC) for each model to balance error scores with model complexity. We assume Gaussian noise and set $\text{AIC} = 2N + V \log(\frac{\text{RSS}}{V})$, where $N$ is the number of model parameters, $V$ is the number of data points (GM regions), and the RSS is the residual sum of squares between the predicted outcome of
the model and the data, i.e. \( \sum_{v=1}^{V} ((Y - \Sigma \beta)^2)_{v} \). The single-best fitting descriptor choice has just one parameter, with trans-neuronal spread or proximity spread having one additional parameter for epicenter selection. The topological profiles have one parameter for each descriptor, plus one parameter for epicenter selection, minus one parameter as the weights are normalized, i.e. as many parameters as the number of the descriptors. The AIC scores are in Supplementary Information Table S4.

Data availability

AD data set from ADNI. ADNI is a public-private partnership. All ADNI data are shared without embargo through the LONI Image and Data Archive (https://ida.loni.usc.edu/login.jsp) a secure research data repository. Interested scientists may obtain access to ADNI imaging, clinical, genomic, and biomarker data for the purposes of scientific investigation, teaching, or planning clinical research studies. Access is contingent on adherence to the ADNI Data Use Agreement. For up-to-date information please see http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_DSP_Policy.pdf.

PPMS data set from UCLH. Data can be obtained upon request, directed the management team of the data at the Institute of Neurology, UCL: uclh.qsmc@nhs.net.

HA data set from the Rotterdam Study. Data can be obtained upon request. Requests should be directed towards the management team of the Rotterdam Study (secretariat.epi@erasmusmc.nl), which has a protocol for approving data requests. Because of restrictions based on privacy regulations and informed consent of the participants, data cannot be made freely available in a public repository. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). The Rotterdam Study has been entered into the Netherlands National Trial Register (NTR; www.trialregister.nl) and into the WHO International Clinical Trials Registry Platform (ICTRP; www.who.int/ictrp/network/primary/en/) under shared catalogue number NTR6831. All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians.

HCP data are from the Human Connectome Project. Open Access Data (all imaging data and most of the behavioral data) is available to those who register an account at ConnectomeDB and agree to the Open Access Data Use Terms. This includes agreement to comply with institutional rules and regulations. For up-to-date information please see https://www.humanconnectome.org/study/hcp-young-adult/data-use-terms.

AD, PPMS and HA processed, anonymized and unrecognizable data, useful to process the mechanistic weights can be obtained upon request to the corresponding author: sara.garbarino@inria.fr.

Code availability

Code is available at: https://github.com/sgarbarino/mechanistic-profiles.

The software (BrainPainter) for coloring brain images as in Figure 2 is open-source and available at: https://github.com/mrazvan22/brain-coloring.

Acknowledgments

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 666992.
SG, NPO, EJV, OC, FB, JMS, MWV, DCA acknowledge funding from the European Union’s Horizon 2020 research and innovation programme.

SG acknowledges financial support from the French government managed by L’Agence Nationale de la Recherche under Investissements d’Avenir UCAEDR (ANR-15-IDEX-01) through the project “AtroProDem: A data-driven model of mechanistic brain Atrophy Propagation in Dementia”.

NPO acknowledges financial support from The Michael J. Fox Foundation for Parkinson’s Research, the Alzheimer’s Association, Alzheimer’s Research UK, and the Weston Brain Institute through “NetMON: Network Models Of Neurodegeneration” (BAND-15-368107, 11042).

DCA and NPO were also funded from EPSRC grants EP/M020533/1 and EP/J020990/01. DCA, NPO, FB, and JMS are supported by the NIHR UCLH Biomedical Research Centre.

Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research and Development, LLC.; Johnson and Johnson, Pharmaceutical Research and Development LLC.; Lumosity; Lundbeck; Merck and Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuroimaging at the University of Southern California.

The Rotterdam Study is supported by the Erasmus MC and Erasmus University Rotterdam; the Netherlands Organization for Scientific Research; the Netherlands Organization for Health Research and Development; the Research Institute for Diseases in the Elderly; the Netherlands Genomics Initiative; the Ministry of Education, Culture and Science; the Ministry of Health Welfare and Sports; the European Commission; and the Municipality of Rotterdam.

The authors acknowledge The National Institute for Health Research (NIHR) Biomedical Research Centre (BRC) at University College London Hospitals (UCLH) for their support.

Declaration of competing interests

The authors declare no competing interests.

References

1. Iturria-Medina, Y. Anatomical Brain Networks on the Prediction of Abnormal Brain States. Brain Connect. 3, 1–21 (2013).

2. Iturria-Medina, Y. & Evans, A. C. On the central role of brain connectivity in neurodegenerative disease progression. Front. Aging Neurosci. 7, 1–10 (2015).
3. Raj, A., Kuceyeski, A. & Weiner, M. A Network Diffusion Model of Disease Progression in Dementia. *Neuron* **73**, 1204–1215 (2012).

4. Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L. & Michael, D. Neurodegenerative diseases target large-scale human brain networks. *Neuron* **62**, 42–52 (2009).

5. Zhou, J., Gennatas, E. D., Kramer, J. H., Miller, B. L. & Seeley, W. W. Predicting Regional Neurodegeneration from the Healthy Brain Functional Connectome. *Neuron* **73**, 1216–1227 (2012).

6. Soto, C. & Pritzkow, S. Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nat. Neurosci.* **21**, 1332–1340 (2018).

7. Jucker, M. & Walker, L. C. Propagation and spread of pathogenic protein assemblies in neurodegenerative diseases. *Nat. Neurosci.* **21**, 1341–1349 (2018).

8. Fu, H., Hardy, J. & Duff, K. E. Selective vulnerability in neurodegenerative diseases. *Nat. Neurosci.* **21**, 1350–1358 (2018).

9. Buckner, R. L. *et al.* Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer’s disease. *J. Neurosci.* **29**, 1860–73 (2009).

10. Saxena, S. & Caroni, P. Selective Neuronal Vulnerability in Neurodegenerative Diseases: from Stressor Thresholds to Degeneration. *Neuron* **71**, 35–48 (2011).

11. Appel, S. H. A unifying hypothesis for the cause of amyotrophic lateral sclerosis, parkinsonism, and alzheimer disease. *Ann. Neurol.* **10**, 499–505 (1981).

12. Salehi, A. *et al.* Increased App Expression in a Mouse Model of Down’s Syndrome Disrupts NGF Transport and Causes Cholinergic Neuron Degeneration. *Neuron* **51**, 29–42 (2006).

13. Frost, B. & Diamond, M. I. Prion-like mechanisms in neurodegenerative diseases. *Nature Reviews Neuroscience* **11**, 155–159 (2010).

14. Jucker, M. & Walker, L. C. Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature* **501**, 45–51 (2013).

15. Prusiner, S. B. Some Speculations about Prions, Amyloid, and Alzheimer’s Disease. *N. Engl. J. Med.* **310**, 661–663 (1984).

16. Buckner, R. L. *et al.* Molecular, structural, and functional characterization of Alzheimer’s disease: evidence for a relationship between default activity, amyloid, and memory. *J. Neurosci.* **25**, 7709–17 (2005).

17. Cope, T. E. *et al.* Tau burden and the functional connectome in Alzheimer’s disease and progressive supranuclear palsy. *Brain* **141**, 550–567 (2018).

18. Fornito, A., Zalesky, A. & Breakspear, M. The connectomics of brain disorders. *Nat. Rev. Neurosci.* **16**, 159–172 (2015).

19. Mancini, M. *et al.* Network attack simulations in Alzheimer’s disease: The link between network tolerance and neurodegeneration. in 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI) 237–240 (IEEE, 2016). doi:10.1109/ISBI.2016.7493253

20. Raj, A. *et al.* Network Diffusion Model of Progression Predicts Longitudinal Patterns of Atrophy and Metabolism in Alzheimer’s Disease. *Cell Rep.* **10**, 359–369 (2015).

21. Raj, A. *et al.* A Network Diffusion Model of Disease Progression in Dementia. *Neuron* **73**, 1204–1215 (2013).
22. Iaccarino, L. et al. Local and distant relationships between amyloid, tau and neurodegeneration in Alzheimer’s Disease. *NeuroImage Clin.* **17**, 452–464 (2018).

23. Weickenmeier, J., Jucker, M., Goriely, A. & Kuhl, E. A physics-based model explains the prion-like features of neurodegeneration in Alzheimer’s disease, Parkinson’s disease, and amyotrophic lateral sclerosis. *J. Mech. Phys. Solids* **124**, 264–281 (2019).

24. Mišić, B. et al. Cooperative and Competitive Spreading Dynamics on the Human Connectome. *Neuron* **86**, 1518–1529 (2015).

25. Iturria-Medina, Y., Carbonell, F. M., Sotero, R. C., Chouinard-Decorte, F. & Evans, A. C. Multifactorial causal model of brain (dis)organization and therapeutic intervention: Application to Alzheimer’s disease. *Neuroimage* **152**, 60–77 (2017).

26. Iturria-Medina, Y., Carbonell, F. M. & Evans, A. C. Multimodal imaging-based therapeutic fingerprints for optimizing personalized interventions: Application to neurodegeneration. *Neuroimage* **179**, 40–50 (2018).

27. Jones, D. T. et al. Cascading network failure across the Alzheimer’s disease spectrum. *Brain* **139**, 547–562 (2016).

28. Leal, S. L., Lockhart, S. N., Maass, A., Bell, R. K. & Jagust, W. J. Subthreshold Amyloid Predicts Tau Deposition in Aging. *J. Neurosci.* **38**, 4482–4489 (2018).

29. Steenwijk, M. D. et al. Cortical atrophy patterns in multiple sclerosis are non-random and clinically relevant. *Brain* **139**, 115–126 (2016).

30. Irwin, D. J., Lee, V. M. Y. & Trojanowski, J. Q. Parkinson’s disease dementia: Convergence of α-synuclein, tau and amyloid-β pathologies. *Nature Reviews Neuroscience* **14**, 626–636 (2013).

31. Lei, P. et al. Tau protein: Relevance to Parkinson’s disease. *International Journal of Biochemistry and Cell Biology* **42**, 1775–1778 (2010).

32. Arai, T. et al. Phosphorylated TDP-43 in Alzheimer’s disease and dementia with Lewy bodies. *Acta Neuropathol.* **117**, 125–136 (2009).

33. Hamilton, R. L. Lewy Bodies in Alzheimer’s Disease: A Neuropathological Review of 145 Cases Using α-Synuclein Immunohistochemistry. *Brain Pathol.* **10**, 378–384 (2006).

34. Spires-Jones, T. L., Attems, J. & Thal, D. R. Interactions of pathological proteins in neurodegenerative diseases. *Acta Neuropathol.* **134**, 187–205 (2017).

35. Attems, J. & Jellinger, K. A. The overlap between vascular disease and Alzheimer’s disease - lessons from pathology. *BMC Med.* **12**, 206 (2014).

36. Kalaria, R. N. Neurodegenerative disease: Diabetes, microvascular pathology and Alzheimer disease. *Nat. Rev. Neurol.* **5**, 305–6 (2009).

37. Sweeney, M. D., Kisler, K., Montagne, A., Toga, A. W. & Zlokovic, B. V. The role of brain vasculature in neurodegenerative disorders. *Nat. Neurosci.* **21**, 1318–1331 (2018).

38. Trapp, B. D. & Stys, P. K. Virtual hypoxia and chronic necrosis of demyelinated axons in multiple sclerosis. *Lancet Neurol.* **8**, 280–291 (2009).

39. Oxtoby, N. P. & Alexander, D. C. Imaging plus X. *Curr. Opin. Neurol.* **30**, 371–379 (2017).

40. Bilgel, M., Prince, J. L., Wong, D. F., Resnick, S. M. & Jedynak, B. M. A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. *Neuroimage* **21**, 205–221 (2018).
41. Donohue, M. C. et al. Estimating long-term multivariate progression from short-term data. *Alzheimer’s Dement.* **10**, S400–S410 (2014).

42. Eshaghi, A. et al. Progression of regional grey matter atrophy in multiple sclerosis. *Brain* **141**, 1665–1677 (2018).

43. Firth, N. C. et al. Data-driven disease progression modelling using neurophysiological tests Posterior Cortical Atrophy vs Alzheimer’s disease. *Alzheimer’s Dement.* **12**, P963–P964 (2016).

44. Fonteijn, H. M. et al. An event-based model for disease progression and its application in familial Alzheimer’s disease and Huntington’s disease. *Neuroimage* **60**, 1880–1889 (2012).

45. Iturria-Medina, Y. et al. Early role of vascular dysregulation on late-onset Alzheimer’s disease based on multifactorial data-driven analysis. *Nat. Commun.* **7**, (2016).

46. Jedynak, B. M. et al. A computational neurodegenerative disease progression score: Method and results with the Alzheimer’s disease neuroimaging initiative cohort. *Neuroimage* **63**, 1478–1486 (2012).

47. Koval, I. et al. Statistical learning of spatiotemporal patterns from longitudinal manifold-valued networks. in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **10433 LNCS**, 451–459 (Springer, Cham, 2017).

48. Li, D., Iddi, S., Thompson, W. K. & Donohue, M. C. Bayesian latent time joint mixed effect models for multicohort longitudinal data. *Statistical Methods in Medical Research* 096228021773756 (2017). doi:10.1177/0962280217737566

49. Marinescu, R. V. et al. DIVE: A spatiotemporal progression model of brain pathology in neurodegenerative disorders. *Neuroimage* **192**, 166–177 (2019).

50. Oxtoby, N. P. et al. Data-driven sequence of changes to anatomical brain connectivity in sporadic Alzheimer’s disease. *Front. Neurol.* **8**, (2017).

51. Oxtoby, N. P. et al. Data-driven models of dominantly-inherited Alzheimer’s disease progression. *Brain* **141**, 1529–1544 (2018).

52. Schiratti, J.-B., Allassonnière, S., Colliot, O. & Durrleman, S. A Bayesian Mixed-Effects Model to Learn Trajectories of Changes from Repeated Manifold-Valued Observations. *J. Mach. Learn. Res.* **18**, 1–33 (2017).

53. Venkatraghavan, V., Bron, E., Niessen, W. & Klein, S. A Discriminative Event Based Model for Alzheimer’s Disease Progression Modeling. 121–133 (2017).

54. Young, A. L. et al. A data-driven model of biomarker changes in sporadic Alzheimer’s disease. *Brain* **137**, 2564–2577 (2014).

55. Hinrichs, C., Singh, V., Xu, G. & Johnson, S. C. Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. *Neuroimage* **55**, 574–589 (2011).

56. Moradi, E., Pepe, A., Gaser, C., Huttunen, H. & Tohka, J. Machine learning framework for early MRI-based Alzheimer’s conversion prediction in MCI subjects. *Neuroimage* **104**, 398–412 (2015).

57. Nie, L. et al. Modeling Disease Progression via Multisource Multitask Learners: A Case Study With Alzheimer’s Disease. *IEEE Trans. Neural Networks Learn. Syst.* **28**, 1508–1519 (2017).
59. Lorenzi, M. et al. Probabilistic disease progression modeling to characterize diagnostic uncertainty: Application to staging and prediction in Alzheimer’s disease. *Neuroimage* **190**, 56–68 (2019).

60. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).

61. Braak, H. & Braak, E. Staging of Alzheimer’s disease-related neurofibrillary changes. *Neurobiol. Aging* **16**, 271–8; discussion 278-84 (1995).

62. Eshaghi, A. et al. Deep gray matter volume loss drives disability worsening in multiple sclerosis. *Ann. Neurol.* **83**, 210–222 (2018).

63. Vinke, E. J. et al. Trajectories of imaging markers in brain aging: the Rotterdam Study. *Neurobiol. Aging* **71**, 32–40 (2018).

64. Narvacan, K., Treit, S., Camicioli, R., Martin, W. & Beaulieu, C. Evolution of deep gray matter volume across the human lifespan. *Hum. Brain Mapp.* **38**, 3771–3790 (2017).

65. Watson, R., Colloby, S. J., Blamire, A. M. & O’Brien, J. T. Subcortical volume changes in dementia with Lewy bodies and Alzheimer’s disease. A comparison with healthy aging. *Int. Psychogeriatrics* **28**, 529–536 (2016).

66. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).

67. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).

68. Zhou, J., Gennatas, E. D., Kramer, J. H., Miller, B. L. & Seeley, W. W. Predicting Regional Neodegeneration from the Healthy Brain Functional Connectome. *Neuron* **73**, 1216–1227 (2012).

69. Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L. & Greicius, M. D. Neurodegenerative Diseases Target Large-Scale Human Brain Networks. *Neuron* **62**, 42–52 (2009).

70. de Haan, W., Mott, K., van Straaten, E. C. W., Scheltens, P. & Stam, C. J. Activity Dependent Degeneration Explains Hub Vulnerability in Alzheimer’s Disease. *PLoS Comput. Biol.* **8**, e1002582 (2012).

71. Alladi, S. et al. Focal cortical presentations of Alzheimer’s disease. *Brain* **130**, 2636–2645 (2007).

72. Du, A.-T. et al. Different regional patterns of cortical thinning in Alzheimer’s disease and frontotemporal dementia. *Brain* **130**, 1159–1166 (2006).

73. Mendez, M. F., Ghajariania, M. & Perryman, K. M. Posterior cortical atrophy: clinical characteristics and differences compared to Alzheimer’s disease. *Dement. Geriatr. Cogn. Disord.* **14**, 33–40 (2002).

74. Barnham, K. J., Masters, C. L. & Bush, A. I. Neurodegenerative diseases and oxidative stress. *Nat. Rev. Drug Discov.* **3**, 205–214 (2004).

75. Cagnin, A. et al. In-vivo measurement of activated microglia in dementia. *Lancet* **358**, 461–467 (2001).

76. Hickman, S., Izzy, S., Sen, P., Morsett, L. & El Khoury, J. Microglia in neurodegeneration. *Nat. Neurosci.* **21**, 1359–1369 (2018).

77. Chiaravalloti, N. D. & DeLuca, J. Cognitive impairment in multiple sclerosis. *Lancet Neurol.* **7**, 1139–1151 (2008).
78. Maier-Hein, K. H. et al. The challenge of mapping the human connectome based on diffusion tractography. *Nat. Commun.* **8**, 1349 (2017).

79. Thomas, C. et al. Anatomical accuracy of brain connections derived from diffusion MRI tractography is inherently limited. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 16574–9 (2014).

80. Petersen, M. V. et al. Probabilistic versus deterministic tractography for delineation of the cortico-subthalamic hyperdirect pathway in patients with Parkinson disease selected for deep brain stimulation. *J. Neurosurg.* **126**, 1657–1668 (2017).

81. Descoteaux, M., Deriche, R., Knosche, T. R. & Anwander, A. Deterministic and Probabilistic Tractography Based on Complex Fibre Orientation Distributions. *IEEE Trans. Med. Imaging* **28**, 269–286 (2009).

82. Tsai, S.-Y. Reproducibility of structural brain connectivity and network metrics using probabilistic diffusion tractography. *Sci. Rep.* **8**, 11562 (2018).

83. Daducci, A., Dal Palù, A., Descoteaux, M. & Thiran, J.-P. Microstructure Informed Tractography: Pitfalls and Open Challenges. *Front. Neurosci.* **10**, 247 (2016).

84. Fornari, S., Schafer, A., Jucker, M., Goriely, A. & Kuhl, E. Prion-like spreading of Alzheimer’s disease within the brain’s connectome. bioRxiv 529438 (2019). doi:10.1101/529438

85. Zeighami, Y. et al. Network structure of brain atrophy in de novo parkinson’s disease. *Elife* **4**, 1–20 (2015).

86. Powell, F. et al. Preserved Structural Network Organization Mediates Pathology Spread in Alzheimer’s Disease Spectrum Despite Loss of White Matter Tract Integrity. *J. Alzheimer’s Dis.* **65**, 747–764 (2018).

87. Acosta, D., Powell, F., Zhao, Y. & Raj, A. Regional vulnerability in Alzheimer’s disease: The role of cell-autonomous and transneuronal processes. *Alzheimers. Dement.* **14**, 797–810 (2018).

88. Young, A. L. et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference. *Nat. Commun.* **9**, 4273 (2018).

89. Mezias, C. & Raj, A. Analysis of Amyloid-β Pathology Spread in Mouse Models Suggests Spread Is Driven by Spatial Proximity, Not Connectivity. *Front. Neurol.* **8**, 653 (2017).

90. Eshaghi, A. et al. Temporal and spatial evolution of grey matter atrophy in primary progressive multiple sclerosis. *Neuroimage* **86**, 257–264 (2014).

91. Hofman, A. et al. The Rotterdam Study: 2016 objectives and design update. *Eur. J. Epidemiol.* **30**, 661–708 (2015).

92. Ikram, M. A. et al. The Rotterdam Scan Study: design update 2016 and main findings. *Eur. J. Epidemiol.* **30**, 1299–1315 (2015).

93. Veraart, J., Fieremans, E. & Novikov, D. S. Diffusion MRI noise mapping using random matrix theory. Magn. Reson. Med. **00**, n/a-n/a (2015).

94. Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* **125**, 1063–1078 (2016).

95. Tustison, N. J. et al. N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).

96. Modat, M. et al. Global image registration using a symmetric block-matching approach. *J. Med.*
Figure legends

Figure 1. Overview of topological profile estimation. (a) we construct the average structural connectome from Human Connectome Project (HCP) tractograms; (b) we compute topological descriptors on the structural connectome and the progression pattern that corresponds to each; (c) we estimate the long-term atrophy progression pattern and its variability within each condition, using GP Disease Progression Model on regional volumes from T1-weighted MRI; (d) we estimate rates of progression for each individual from
the cohort-level GP progression model; (e) we estimate each topological profile (both cohort-level and individual) as the linear combination of topological descriptors, with weights $\beta$, that best matches the observed progression rates. Those profiles are then visualized in a low-dimensional projection of the space of topological descriptors.

Figure 2. Temporal evolution of brain loss in AD, PPMS and HA confirm known atrophy progression patterns, and the progression patterns corresponding to the topological profiles for the three cohorts match the progression of atrophy better than the single best fitting topological descriptor. Top row: 4D representation of the GP disease progression model for AD (left), PPMS (middle) and HA (right). Second row: 4D representation of the progression pattern corresponding to the topological profile for AD (left) - a combination of centrality measures and network-based proximity; PPMS (middle) – a combination of centrality, segregation and cortical proximity measures; and HA (right) - a combination of centrality, cortical proximity, and constant progression. Third row: 4D representation of the progression pattern corresponding the single best fitting topological descriptor for AD (left) - network proximity; PPMS (middle) - segregation; and HA (right) - cortical proximity. Each region’s color opacity is proportional to the cumulative abnormality of each region (strong blue means strongly atrophied), and time increases from left to right. AIC is the Akaike Information Criterion for the fit to the observed disease progression (lower is better).

Figure 3. Individual profiles are specific for each neurological condition. Red indicates AD individuals, blue indicates HA and green is PPMS. Panel (a) shows ternary plot of the individual profiles, obtained via the GP Progression Model, for AD+MCI-diagnosed individuals, PPMS-diagnosed individuals, and HA individuals plotted according to the distance from the cohort-level profile; corners are cohort-level profiles. Outliers of each cohort are highlighted (identified with diamonds); (b) shows a 2D representation of the topological profiles in (a), using tSNE; big stars represent the cohort profiles and small ones the bootstrapped cohort-profiles; (c) ternary plot using GP Progression Model -driven profiles for only healthy control individuals in the AD and PPMS cohorts, and HA individuals; (d) tSNE plot of data in (c); (e) ternary plot of the individual profiles for AD+MCI-diagnosed, PPMS-diagnosed, and HA individuals, estimated from only late-stage data; (f) tSNE plot for the topological profiles of data in (e).

Figure 4: Temporal progression patterns (left-to-right) of different descriptors. For each descriptor (row), abnormality increases in a descriptor-specific pattern. The magnitude of cumulative abnormality at a node is proportional to the color intensity. Red nodes are epicenters. (a) Centrality: node A is affected first due to having the highest centrality, followed by node B, then C and D. (b) Segregation: node D is affected first due to having the highest segregation, followed by C, then B and A. (c) Network proximity: nodes B and C are affected before D, because they are closer to the epicenter A (along the connectivity network). (d) Cortical proximity: node B is affected first because of its spatial proximity to the epicenter A, then C and finally D. Here edges are dashed as no information is needed from connectivity.

Figure 5: A schematic representation of the mathematical modeling of the topological profiling with GP Progression Model. In the example here we have three biomarkers/regions (represented in red ($\nu = 1$), green ($\nu = 2$) and blue ($\nu = 3$)), and two topological descriptors ($g$ and $h$). a) The GP Progression Model estimates temporal trajectories of biomarkers progression, along the disease time $\tau$. The unique maximum points of the derivatives of the trajectories correspond to their maximal rate of change $Y$. b) Two topological descriptors are computed for each region from anatomical connectomes. They combine, column-wise, in the matrix $Z$. c) For each subject $j$, the GP Progression Model estimates a time-reparametrization which shifts individual measurements to the disease time. For each biomarker, the speed of progression of subject $j$ with respect to the cohort progression is the value of the derivative of the biomarker progression at $\tau$ $j$, which represents the shift of the average age of the subject. d) Topological profiles are estimated via a linear model relating $Z$ and $Y$ (for the cohort-level topological profile) or $Y_j$ (for the individual profiles).
Supplementary Figure legends

Figure 2–figure supplement 1. Fine-grained representation of Figure 2, with 12 stages. Top row, for each panel: 4D representation of the disease progression model for AD (top panel), PPMS (middle panel) and HA (bottom panel). Middle row, for each panel: 4D representation of the progression corresponding to the topological profiles for AD (top panel), combination centrality and network proximity; PPMS (middle panel), combination of segregation, centrality and cortical proximity; and HA (bottom panel), combination of centrality and cortical proximity spread. Bottom row, for each panel: 4D representation of the progression corresponding to the single best fitting descriptors for AD (top panel), network proximity to epicenter; PPMS (middle panel), segregation - inverse clustering specifically; HA (bottom panel), cortical proximity. Each region’s color is proportional to the cumulative abnormality of each region, and time reads from left to right. In red rectangles, the regions selected for Fig 2, main text.

Figure 2–figure supplement 2. Biomarker trajectories, with standard deviations and measurements for the AD cohort. Here in full red the estimated trajectories and in dashed red the standard deviations (1-sigma confidence band). Each colored dot is a individual, and is colored according to the baseline diagnosis (magenta = HC, green = MCI, black = AD). The model estimates a time-shift (“Re-parameterized age”) that shuffles the individuals disaggregating the diagnosis (i.e. HC individual are earlier in the progression, followed by MCI, and then AD).

Figure 2–figure supplement 3. Individual trajectories, with standard deviations and measurements, for the PPMS cohort. Estimated trajectories in full red and in dashed red the standard deviations (1-sigma confidence band). Each colored dot is a individual, and is colored according to the baseline diagnosis (magenta = HC, green = PPMS). The model estimates a time-shift (“Re-parameterized age”) that shuffles the individuals disaggregating the diagnosis (i.e. HC individual are earlier in the progression, followed by PPMS).

Figure 2–figure supplement 4. Individual trajectories, with standard deviations and measurements, for the HA cohort. Estimated trajectories in full red and in dashed red the standard deviations (1-sigma confidence band). Each colored dot is a individual, and is colored according to the baseline age: magenta < mean-std (Young” cohort); mean-std<green<mean+std; black > mean+std (“Old” cohort).

Figure 3–figure supplement 1. Longitudinal information for the study cohort (AD: 1713 individuals, HA 5463 individuals, PPMS 64 individuals). The overall time-points are 6670 for AD, 11627 for HA and 244 for PPMS; the average follow-up (years) are 2.4 for AD, 5.3 for HA and 4.0 for PPMS; the standard deviation of follow-up (years) is 1.8 for AD, 1.1 for HA and 1.5 for PPMS; the range of follow-up (years) is 0.5-10 for AD, 0.7-10.5 for HA and 0-6 for PPMS.

Supplementary File 1. Supplementary file comprising Tables S1-S9.
Figure 1

(a) HCP cohort Tractograms
   Structural connectomes

(b) Topological descriptor progression
   Centrality
   Segregation

(c) AD, PPMS or HA cohort
   T1-MRI segmentation
   Cohort-level DPM progression

(d) Individual trajectories

(e) Cohort topological profile

\[ \text{cohort DPM} \sim \text{linear}(\beta_{\text{centrality}}, \beta_{\text{segregation}}) \]

individual trajectories \( \sim \) linear(\( \beta_{\text{centrality}}, \beta_{\text{segregation}} \))
| Disease progression | AD            | PPMS          | HA            |
|---------------------|---------------|---------------|---------------|
| Topological profile | ![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) |
| AIC = 24.4          | AIC = 24.8    | AIC = 20.1    |
| Strongest descriptor| ![Image](image4.png) | ![Image](image5.png) | ![Image](image6.png) |
| AIC = 35.9          | AIC = 31.0    | AIC = 27.6    |
Figure 3

GP progression model-driven profiles (patients)

(a) AD, HA, PPMS

GP progression model-driven profiles (controls)

c (HA, AD, PPMS)

d

Late-stage-driven profiles (patients)

e (HA, AD, PPMS)

f
Figure 5

(a) Biomarkers’ measurements vs. \( \tau \) with rate of change.

(b) Topological descriptor progression for descriptors \( g \) and \( h \).

(c) Individualization procedure with individual topological profiles.

(d) Cohort-level topological profiles.

GP Progression Model – cohort-level progression:

\[
Y = \Sigma \beta = \begin{bmatrix} g_1 & h_1 \\ g_2 & h_2 \\ g_3 & h_3 \end{bmatrix}
\]

\[
Y^j = \Sigma \beta^j = \begin{bmatrix} Y_2^j & Y_1^j & Y_3^j \end{bmatrix}^T
\]

\[
\beta = (\beta_g, \beta_h)^T
\]

\[
\beta^j = (\beta_g^j, \beta_h^j)^T
\]
Figure 2—figure supplement 3

Abnormality vs. Re-parametrized Age for various brain regions:
- bankssts
- caudal ant. cing.
- caudal mid. front.
- cuneus
- entorhinal
- fusiform
- inf. parietal
- inthms cing.
- lat. occ.
- lat. orbitofront.
- lingual
- medial orbitofront.
- pars operc.
- pars orb.
- pars triang.
- pericalcarine
- postcentral
- precenral
- precuneus
- rostral ant. cing.
- rostral mid. front.
- sup. front.
- sup. par
- temp. pole
- transverse temp.
- insula
- thalamus
- sup. temp.
- supramarginal
- front. pole
- pallidum
- hippo
- amygdala
- accumbens
