Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# A Wasserstein Weisfeiler-Lehman Kernel Framework for Multimodal Brain Connectome Analysis

**Chiara Razzetta**

chiara.razzetta@aomliguria.it

IRCCS Azienda Ospedaliera Metropolitana

**Sara Garbarino**

University of Genoa

**Research Article**

**Keywords:**

**Posted Date:** October 29th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11266269/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

# A Wasserstein Weisfeiler-Lehman Kernel Framework for Multimodal Brain Connectome Analysis

Chiara Razzetta[a,*], Sara Garbarino[a,b]

[a] *LISCOMP Lab, AOM IRCCS Ospedale Policlinico San Martino, Genova, Italy*
[b] *DIMA, Università degli Studi di Genova, Genova, Italy*

**Abstract**

We introduce a Wasserstein Weisfeiler-Lehman (WWL) graph-kernel framework for the joint analysis of weighted structural graphs and continuous node attributes, together with a regularized fractional-Laplacian extension of the underlying graph propagation operator. The proposed fractional operator provides non-local propagation across the entire graph, with influence decaying as a power law with graph distance, while remaining anchored to the subject-specific structural topology. This extends the standard finite-range Weisfeiler Lehman propagation scheme to a controlled long-range setting without introducing an externally prescribed coupling between structural and functional information.

We formulate the framework for multimodal brain connectivity by representing each subject as a single attributed graph: diffusion tractography defines the weighted structural edges, while resting-state functional connectivity profiles are encoded as continuous node attributes. We evaluate the method on a synthetic cohort of 900 subjects in three ordered clinical groups along the Alzheimer's disease spectrum, generated from multiple structural templates and characterised by a spatially structured loss of resting-state network segregation with age. Under nested cross-validation, the WWL kernel achieves a balanced accuracy of $0.651 \pm 0.028$, significantly outperforming both a raw-feature SVM and a histogram-based WL-kernel baseline.

Controlled ablations separate the contribution of the Wasserstein representation from that of structural propagation. Replacing the propagation operator by the identity yields essentially identical classification perfor-

*Corresponding author
*Email address:* chiara.razzetta@aomliguria.it (Chiara Razzetta)

mance, whereas a structure-only representation approaches chance, showing that the discriminative gain in this benchmark is primarily associated with the Wasserstein point-cloud representation rather than structural propagation. Conversely, experiments across distinct signal regimes show that the fractional extension is signal-dependent: non-local propagation can improve continuous prediction for diffuse whole-brain coupling signals, but provides no consistent advantage for network-segregation signals. We further identify a connectivity artefact induced by fixed-percentile structural sparsification and establish a threshold preserving graph connectivity across subjects. The results clarify the respective roles of Wasserstein geometry, structural propagation, and non-locality in multimodal graph representations and provide a controlled framework for assessing when fractional graph propagation is expected to be beneficial.

*Keywords:* Graph kernels, Wasserstein distance, Weisfeiler-Lehman, Fractional graph Laplacian, Multimodal brain connectivity, Alzheimer's disease

## 1. Introduction

The human brain can naturally be represented as a network, with brain regions serving as nodes and anatomical or statistical relationships between them defining edges. Neuroimaging provides complementary measurements of these relationships. Diffusion tensor imaging (DTI) characterises white-matter organisation through measures such as fractional anisotropy (FA) or streamline counts, yielding a structural connectome, whereas resting-state functional magnetic resonance imaging (rs-fMRI) measures statistical dependencies between regional BOLD signals, yielding a functional connectome. Structural and functional connectivity are therefore not independent descriptions: the anatomical scaffold constrains and shapes functional dynamics [10, 19]. Their joint analysis consequently requires a representation capable of retaining both the topology of the structural network and the spatial organisation of functional information.

A common mathematical framework for combining these modalities is provided by multilayer or multiplex networks, in which structural and functional connectomes are represented as distinct layers defined over a shared set of nodes [3, 6, 8]. Interlayer edges are then introduced to quantify relationships between the two layers, for example through correlations between

structural and functional connectivity profiles [2], and multimodal information is analysed through multilayer graph measures. Such approaches have been applied to several neurological and psychiatric conditions, including multiple sclerosis, depression, and Alzheimer's disease [1, 9]. From a mathematical perspective, however, the structural and functional networks remain separate objects: their interaction is mediated by an additional set of interlayer weights whose definition is external to the propagation dynamics within either layer.

Graph kernels provide an alternative formulation in which multimodal information can be encoded in a single graph and subjects can be compared directly through a positive-semidefinite similarity function. This avoids the need to construct separate feature spaces or to specify an external coupling between modalities [12, 14]. In particular, the Wasserstein Weisfeiler-Lehman (WWL) kernel [20] extends the classical Weisfeiler-Lehman framework by representing neighbourhood information through probability measures and comparing the resulting node representations using Wasserstein distances. This construction is particularly suitable for attributed graphs with continuous-valued node information, since it retains the geometry of the underlying feature distributions rather than reducing them to discrete labels or histogram counts.

We formulate multimodal brain connectivity within this framework by assigning distinct mathematical roles to structural and functional information. Each subject is represented by a single weighted graph whose edges are determined by the structural connectome, while functional connectivity profiles are associated with the nodes as continuous attributes. Structural connectivity therefore defines the topology on which the Weisfeiler-Lehman propagation operates, whereas functional connectivity constitutes the signal propagated through that topology. The interaction between the two modalities is thus encoded directly in the graph propagation operator rather than through a separately estimated interlayer coupling. Iterative propagation produces a sequence of multi-scale node representations, and the resulting point clouds are compared between subjects through Wasserstein geometry, yielding a kernel matrix that can be used directly by kernel-based learning algorithms.

The standard propagation scheme is local: after $H$ iterations, information can only travel along structural paths of at most $H$ edges. This finite propagation range is natural for the classical Weisfeiler-Lehman construction, but it imposes a hard spatial cutoff. In a brain network, however, regions sepa-

rated by many structural hops may still participate in coordinated functional processes. This motivates the question of whether structural information can be propagated over the entire graph while retaining a controlled dependence on graph distance.

To this end, we introduce a regularized fractional-Laplacian generalisation of the propagation operator, following Filippo and Mazza [16]. Fractional powers of graph Laplacians provide a natural mechanism for non-local diffusion, allowing information to reach arbitrarily distant nodes while assigning progressively smaller influence to longer graph paths. The regularisation ensures that the resulting operator remains well behaved and numerically stable. In the present setting, the fractional operator therefore provides a mathematically controlled extension from finite-range Weisfeiler-Lehman propagation to long-range graph interactions, while preserving the subject-specific structural topology.

The introduction of this non-local operator also raises a more general methodological question: when does structural propagation actually contribute to the information captured by the kernel? A gain in classification performance alone cannot distinguish between information originating from the graph topology, information already present in the functional node attributes, and the geometry induced by the Wasserstein representation. Moreover, the usefulness of non-local propagation may depend on the spatial scale of the underlying signal. A diffuse whole-brain alteration may benefit from long-range propagation, whereas a signal concentrated within specific functional systems could instead be diluted by interactions with distant regions.

We address these questions through a controlled synthetic framework in which structural and functional components can be manipulated independently. Rather than considering a binary disease contrast characterised by a global change in connectivity magnitude, we construct a cohort of $S = 900$ subjects in three ordered clinical groups, pooled from six independently seeded structural templates. The functional signal is designed as a progressive loss of resting-state network segregation with age, thereby introducing a spatially structured, network-level alteration that cannot be represented simply as a uniform rescaling of whole-brain connectivity. This setting allows us to evaluate both the discriminative properties of the WWL representation and the additional contribution of structural propagation.

We compare the standard local propagation operator with the proposed regularized fractional operator and with representations designed to isolate individual components of the framework. In particular, we consider a raw-

feature SVM baseline, a histogram-based WL-kernel construction, and an identity-propagation ablation in which structural propagation is removed while retaining the Wasserstein kernel formulation. A structure-only control is also considered. This decomposition allows the role of Wasserstein geometry to be distinguished from that of structural propagation rather than attributing the entire performance of the multimodal representation to the latter.

The experiments reveal two complementary aspects of the framework. First, the Wasserstein point-cloud representation provides the main discriminative contribution for the network-segregation signal considered here, whereas structural propagation does not produce a measurable additional gain. Second, the fractional extension exhibits a signal-dependent behaviour across independent synthetic settings: non-local propagation can improve a continuous age-related readout when the underlying coupling signal is spatially diffuse, while it can be neutral or detrimental when the signal is organised around specific functional networks. Thus, the fractional operator is not treated as a universally beneficial replacement for local propagation, but as a mathematically controlled mechanism whose usefulness depends on the spatial structure of the information to be propagated.

Finally, we identify a methodological issue preceding the kernel construction itself. Fixed-percentile sparsification of structural connectivity, commonly used to obtain comparable graph densities across subjects, can result in disconnected subject-specific graphs. We show that this effect is present both in the synthetic cohort considered here and we determine a lower sparsification threshold that preserves graph connectivity across subjects. This correction is important for propagation-based graph methods because disconnected components impose an artificial constraint on the flow of structural information.

Taken together, these results provide a unified framework for studying multimodal graph representations in which Wasserstein geometry, structural propagation, and graph non-locality can be investigated as separate mathematical ingredients. The brain-connectivity application provides a controlled test bed in which their respective contributions and limitations can be quantified, with particular emphasis on the conditions under which fractional graph propagation can add information beyond a local Wasserstein Weisfeiler-Lehman representation.

The paper is organised as follows: Section 2 presents the theoretical framework, Section 3 reports experimental results, and Section 4 concludes.

## 2. Methods

### *2.1. Background: WL colour refinement and the WWL kernel*

The Weisfeiler-Lehman (WL) colour-refinement procedure [7] assigns each vertex $v$ a colour encoding its local neighbourhood structure. Starting from an initial labelling $\ell^0$, it iterates

$$\ell^{h+1}(v) = \mathrm{hash}\big(\ell^h(v),\ \mathrm{Sort}\left\{\ell^h(u) : u \in \mathcal{N}(v)\right\}\big)\,, \tag{1}$$

compressing the multiset of neighbour colours into a short label via a globally shared hash table. After $H$ rounds, each vertex has seen its $H$-hop neighbourhood. Classical WL-based kernels count shared colours via an inner product over label histograms, which is efficient but discards the distributional geometry of the resulting embeddings.

When node attributes are continuous vectors $a(v) \in \mathbb{R}^m$, Togninalli et al. [20] replace hashing with a weighted neighbourhood average,

$$a^{h+1}(v) = \tfrac{1}{2}\left(a^h(v) + \tfrac{1}{\deg(v)} \sum_{u \in \mathcal{N}(v)} w(v,u)\, a^h(u)\right), \quad a^0(v) = a(v), \tag{2}$$

where $w(v,u)$ are edge weights and the $\frac{1}{2}$ factor keeps feature scales comparable across iterations. After $H$ steps the features at each depth are stacked to form the graph embedding

$$f^H : G \longmapsto \mathrm{concat}\big(X_G^0,\, X_G^1,\, \ldots,\, X_G^H\big)\,, \qquad X_G^h = \big[a^h(v_1), \ldots, a^h(v_n)\big]^\top, \tag{3}$$

representing each graph as a *point cloud* of $n$ vectors in $\mathbb{R}^{m(H+1)}$.

Rather than comparing point clouds via a histogram inner product, the WWL kernel measures the minimum transport cost between them. Given graphs $G$ and $G'$, the $L^1$-Wasserstein distance with Euclidean ground metric is $W_1(X, X') = \min_{P \in \Gamma} \langle P, M \rangle$, where $M_{ij} = \|x_i - x'_j\|$ and $P$ ranges over transport plans with uniform marginals $\mathbf{1}/n$. The kernel is then the Laplacian of this distance,

$$K_{\mathrm{WWL}}(G, G') = \exp\big(-\lambda\, W_1\big(f^H(G),\, f^H(G')\big)\big)\,, \qquad \lambda > 0. \tag{4}$$

For categorical node features (Hamming ground distance) this construction is positive definite for all $\lambda > 0$ [20]; for continuous attributes positive definiteness is empirically satisfied after input standardisation.

*2.2. Multi-layer connectome graph construction*

Let $\mathcal{P} = \{r_1, \ldots, r_N\}$ be a fixed brain parcellation common to all subjects. For each subject $s$, we construct a single weighted attributed graph

$$G_s = (V, E_s, w_s, a_s^0),$$

where $V = \mathcal{P}$, $E_s$ is the subject-specific edge set, $w_s$ assigns a weight to each structural edge, and $a_s^0(r_i) \in \mathbb{R}^m$ denotes the continuous attribute associated with region $r_i$. The proposed graph-kernel formulation is agnostic to the physical meaning of the edge weights and node attributes. Indeed, we instantiate these two mathematical components using structural and functional neuroimaging measurements.

*Structural graph..* The edge set $E_s$ is derived from DTI tractography: $(r_i, r_j) \in E_s$ whenever a supra-threshold white-matter tract connects the two ROIs. Edges with FA below the 70th percentile of the per-subject distribution are removed to suppress spurious tractography connections. The remaining edge weight is

$$w_s(r_i, r_j) = \mathrm{FA}_s(r_i, r_j).$$

The resulting weighted graph defines the topology on which the propagation operator acts.

*Continuous node attributes..* The initial attribute associated with region $r_i$ is the $i$-th row of the subject's functional connectivity matrix derived from rs-fMRI,

$$a_s^0(r_i) = \big(\mathrm{FC}_s\big)_{i,\cdot} = \Big(\mathrm{corr}\big(\mathrm{BOLD}_s^{r_i}, \mathrm{BOLD}_s^{r_j}\big)\Big)_{j=1}^{N} \in \mathbb{R}^N. \tag{5}$$

FC matrices are z-score normalised per subject, using each subject's own row-wise mean and standard deviation. This removes subject-specific global offsets in the functional-connectivity profiles before propagation and is used consistently throughout the study.

*Remark* 1 (Choice of structural and functional roles). The assignment of DTI to graph structure and rs-fMRI to continuous node attributes is a modelling choice motivated by the biological relationship between the two modalities, rather than a requirement of the proposed mathematical framework. DTI provides a subject-specific anatomical scaffold that constrains the spatial organisation of communication between brain regions, and is therefore used to

define the topology and weights of the propagation graph. Resting-state functional connectivity, in contrast, describes statistical relationships between regional activity patterns and is naturally represented as a continuous signal associated with each region. This motivates treating structural connectivity as the *propagation topology* and functional connectivity as the *propagated signal*. The resulting construction encodes the structure-function interaction through the propagation operator itself, without introducing a separate functional graph layer or an externally prescribed interlayer coupling.

Figure 1 summarises the resulting pipeline, from the multimodal graph construction to the subject-by-subject kernel matrix.

### *2.3. WL propagation on the connectome*

Applying Eq. (2) with DTI weights as the propagation operator yields

$$A_s^{h+1} = \tfrac{1}{2}\big(I + D_s^{-1}W_s\big)A_s^h, \qquad h = 0, \ldots, H-1, \tag{6}$$

where $D_s = \mathrm{diag}(\deg_s(r_1), \ldots, \deg_s(r_N))$ is the diagonal weighted-degree matrix and $W_s \in \mathbb{R}^{N\times N}$ is the structural weight matrix. The feature matrix $A_s^h \in \mathbb{R}^{N\times N}$ stores the regional continuous attributes at hop $h$. Multiplying $A_s^h$ on the left by the diffusion kernel $T_s = \frac{1}{2}(I + D_s^{-1}W_s)$ effectively filters each region's functional connectivity profile along the structural paths, mirroring a non-parametric graph convolution [11].

Each row $a_s^h(r_i) \in \mathbb{R}^N$ has a concrete multi-scale interpretation: at depth $h = 0$ it is the raw functional correlation profile of region $r_i$; at depth $h > 0$ it is that profile averaged over the $h$-hop structural neighbourhood, weighted by FA. The functional identity of a region is therefore progressively contextualised by its anatomical surroundings.

After $H$ steps the subject-level embedding is the point cloud $f^H(G_s) = \mathrm{concat}(A_s^0, \ldots, A_s^H) \in \mathbb{R}^{N\times N(H+1)}$, and the connectome WWL kernel is obtained by instantiating Eq. (4):

$$K_{\mathrm{WWL}}(s, s') = \exp\big(-\lambda\, W_1\big(f^H(G_s),\, f^H(G_{s'})\big)\big)\,. \tag{7}$$

Assembling all $\binom{S}{2}$ pairwise values yields the symmetric kernel matrix $\mathbf{K} \in \mathbb{R}^{S\times S}$, where $S$ is the total number of subjects in the population.

*Remark* 2 (Oversmoothing and choice of $H$). The operator $T_s$ has spectral radius 1 [13], so iterating for large $H$ causes node embeddings to converge to the random-walk stationary distribution. In practice $H \in \{1, 2, 3\}$ is sufficient; $H$ is selected per experiment by the inner cross-validation loop over this candidate set (Section 3.2).

*2.4. Non-local structural propagation via a regularized fractional Laplacian*

The propagation operator $T_s = \frac{1}{2}(I + D_s^{-1}W_s)$ in Eq. (6) only ever averages a region's functional profile over its $h$-hop *structural* neighbourhood: after $H$ iterations, information has travelled at most $H$ edges of $G_s$. Two regions joined by a long, indirect white-matter pathway therefore only influence one another once $H$ is large enough to bridge them, and Remark 2 above shows that $H$ cannot be increased indefinitely without oversmoothing every embedding toward the same random-walk stationary point. We introduce an alternative propagation topology that decouples non-local reach from propagation depth, by replacing $W_s$ in Eq. (6) with a structural graph that already encodes long-range reachability, following the regularized fractional graph Laplacian [16].

*Fractional graph Laplacian..* Let $\Delta_s = D_s - W_s$ be the combinatorial structural Laplacian of subject $s$, with spectral decomposition $\Delta_s = Q_s\Lambda_s Q_s^\top$. For $\alpha \in (0, 1)$, the *fractional graph Laplacian* is

$$\Delta_s^\alpha := Q_s\Lambda_s^\alpha Q_s^\top, \tag{8}$$

whose associated *fractional graph* has edge weight $w_s^\alpha(r_i, r_j) = -(\Delta_s^\alpha)_{ij} > 0$ for every pair $i \neq j$, decaying as a power law in the combinatorial hop distance between $r_i$ and $r_j$ on $G_s$ [16]. As $\alpha \to 1^-$, $\Delta_s^\alpha \to \Delta_s$ and the fractional graph collapses back onto $G_s$; as $\alpha \to 0^+$, $\Delta_s^\alpha$ converges (entrywise) to a graph-independent complete graph with uniform weight $1/N$. Thus $\alpha$ interpolates continuously between the purely local operator of Section 2.3 and an entirely non-local, topology-blind one.

*Regularization..* Using $\Delta_s^\alpha$ in place of $\Delta_s$ directly is not, in general, safe: the fractional graph discards every real structural edge weight in favour of the power-law-decayed one, and, as shown in [16], it is only guaranteed to be *superdiffusive* relative to $\Delta_s$, i.e. to actually accelerate structural mixing rather than slow it down, when the subject's largest Laplacian eigenvalue satisfies $\lambda_{\max}(\Delta_s) < 1$, a condition that need not hold for FA-weighted structural connectomes with a small number of high-degree hub regions. We therefore use the *regularized fractional graph* of [16], which keeps every existing structural edge at its own weight and adds the fractional, power-law-decaying weight only where $G_s$ has no direct edge:

$${}_r w_s^\alpha(r_i, r_j) := \begin{cases} w_s(r_i, r_j) & (r_i, r_j) \in E_s, \\ \beta\, w_s^\alpha(r_i, r_j) & \text{otherwise}, \end{cases} \qquad \beta \geq 1, \tag{9}$$

with associated Laplacian ${}_r\Delta_s^\alpha = \Delta_s + \Delta_s^\alpha - \mathrm{restr}(\Delta_s^\alpha)$, where $\mathrm{restr}(\Delta_s^\alpha)$ is the Laplacian of the fractional graph restricted to the edges of $G_s$ [16]. This identity shows that ${}_r\Delta_s^\alpha$ requires no extra asymptotic cost beyond the $O(N^3)$ eigendecomposition of $\Delta_s$ already needed for $\Delta_s^\alpha$, negligible for $N = 100$, and orthogonal to the $O(S^2)$ pairwise Wasserstein cost of Section 2.5. Furthermore, ${}_r\Delta_s^\alpha$ is superdiffusive relative to $\Delta_s$ for *every* $\alpha \in (0, 1)$ and every graph, regardless of $\lambda_{\max}(\Delta_s)$, the property the raw fractional Laplacian only satisfies conditionally.

*Closed form..* Let $B_s \in \{0,1\}^{N\times N}$ be the hollow binary adjacency matrix of $G_s$, $(B_s)_{ij} = \mathbf{1}[(r_i, r_j) \in E_s]$. Eq. (9) admits the single-expression form [16]

$$ {}_rW_s^\alpha = \beta\big(\Delta_s^\alpha \circ (I + B_s) - \Delta_s^\alpha\big) + W_s, \tag{10}$$

here written for general $\beta \geq 1$, which avoids the case distinction of Eq. (9) and is the form used in Algorithm 1 below.

We check the three cases of Eq. (9) entrywise. On the diagonal, $B_s$ is hollow so $(I + B_s)_{ii} = 1$, giving

$$\big(\Delta_s^\alpha \circ (I + B_s) - \Delta_s^\alpha\big)_{ii} = (\Delta_s^\alpha)_{ii} - (\Delta_s^\alpha)_{ii} = 0$$

and hence $({}_rW_s^\alpha)_{ii} = (W_s)_{ii} = 0$, as required of a hollow weight matrix. Off the diagonal with $(r_i, r_j) \in E_s$, i.e. $(B_s)_{ij} = 1$:

$$\big(\Delta_s^\alpha \circ (I + B_s) - \Delta_s^\alpha\big)_{ij} = (\Delta_s^\alpha)_{ij} \cdot (0 + 1) - (\Delta_s^\alpha)_{ij} = 0,$$

so $({}_rW_s^\alpha)_{ij} = \beta \cdot 0 + (W_s)_{ij} = w_s(r_i, r_j)$, matching the first case of Eq. (9). Off the diagonal with $(r_i, r_j) \notin E_s$, i.e. $(B_s)_{ij} = 0$ and hence $(W_s)_{ij} = 0$ (the support of $W_s$ is exactly $E_s$):

$$\big(\Delta_s^\alpha \circ (I + B_s) - \Delta_s^\alpha\big)_{ij} = 0 - (\Delta_s^\alpha)_{ij} = -(\Delta_s^\alpha)_{ij} = w_s^\alpha(r_i, r_j),$$

so $({}_rW_s^\alpha)_{ij} = \beta\, w_s^\alpha(r_i, r_j) + 0$, matching the second case.

*Fractional connectome propagation..* Replacing $W_s$ with ${}_rW_s^\alpha$ in Eq. (6) yields the fractional propagation operator

$$A_s^{h+1} = \tfrac{1}{2}\big(I + ({}_rD_s^\alpha)^{-1}\,{}_rW_s^\alpha\big)A_s^h, \qquad {}_rD_s^\alpha = \mathrm{diag}\big({}_rW_s^\alpha \mathbf{1}\big), \tag{11}$$

used in place of Eq. (6) to build $f^H(G_s)$ and, via Eq. (7), a fractional connectome WWL kernel. Because Eq. (11) already lets a single hop reach every

**Algorithm 1** Regularized fractional WL propagation for one subject

**Require:** structural weight matrix $W_s \in \mathbb{R}^{N\times N}$, initial attributes $A_s^0 \in \mathbb{R}^{N\times N}$, fractional order $\alpha \in (0,1)$, scaling $\beta \geq 1$, depth $H$
**Ensure:** multi-scale embedding $f_\alpha^H(G_s) \in \mathbb{R}^{N\times N(H+1)}$

$B_s \leftarrow (W_s > 0)$ ▷ binary adjacency
$D_s \leftarrow \mathrm{diag}(W_s\mathbf{1})$; $\quad \Delta_s \leftarrow D_s - W_s$ ▷ structural Laplacian
$(Q_s, \Lambda_s) \leftarrow \textsc{eigendecompose}(\Delta_s)$
$(\Lambda_s^\alpha)_{ii} \leftarrow (\Lambda_s)_{ii}^\alpha$ for all $i$ ▷ entrywise fractional power
$\Delta_s^\alpha \leftarrow Q_s\, \Lambda_s^\alpha\, Q_s^\top$
${}_rW_s^\alpha \leftarrow \beta\big(\Delta_s^\alpha \circ (I + B_s) - \Delta_s^\alpha\big) + W_s$ ▷ Eq. (10)
${}_rD_s^\alpha \leftarrow \mathrm{diag}({}_rW_s^\alpha\mathbf{1})$
$T_s^\alpha \leftarrow \frac{1}{2}\big(I + ({}_rD_s^\alpha)^{-1}\,{}_rW_s^\alpha\big)$
**for** $h = 0, \ldots, H-1$ **do**
    $A_s^{h+1} \leftarrow T_s^\alpha\, A_s^h$
**end for**
**return** $\mathrm{concat}\big(A_s^0, \ldots, A_s^H\big)$

region within the subject's own structural graph, it decouples the propagation *depth* $H$, still bounded, for the same oversmoothing reason as Section 2.3, from the propagation *reach*, which $\alpha$ now controls independently. Both $\alpha$ and $\beta$ are treated as additional hyperparameters, selected jointly with $H$, $\lambda$ and $C$ by the same nested cross-validation protocol of Section 3.2.

*Remark* 3 (Reach is bounded by the subject's own connected components). Because $\Delta_s^\alpha$ is a spectral function of $\Delta_s$, it shares $\Delta_s$'s null space: if $G_s$ has $k > 1$ connected components, both $\Delta_s$ and $\Delta_s^\alpha$ have a $k$-dimensional kernel spanned by the components' indicator vectors, so $w_s^\alpha(r_i, r_j) = 0$ whenever $r_i, r_j$ lie in different components. The fractional operator, regularized or not, therefore only ever adds non-local reach *within* each connected component of $G_s$; it cannot substitute for a structural sparsification threshold aggressive enough to fragment the graph in the first place. In practice, the percentile threshold of Section 2.2 should be validated to keep each subject's structural graph in a single connected component before Eq. (11) is used, a condition the fixed 70th-percentile default is not guaranteed to satisfy, and should be checked per cohort rather than assumed.

*2.5. Methodological novelty and biological interpretation*

*Integration vs post-hoc fusion..* Both modalities are integrated within the embedding step itself: the structural topology is the propagation filter and the functional FC profiles are the signal being filtered. This makes the framework sensitive to individual differences in how function is shaped by anatomy. Two subjects with identical marginal FC matrices but different DTI topologies will produce different propagated feature matrices $\{A_s^h\}_{h\geq 1}$ and hence different kernel values, information strictly inaccessible to any single-modality approach or post-hoc feature concatenation.

*Wasserstein distance vs histogram inner product..* Classical WL-based kernels aggregate node colours into label-frequency histograms and compare them via inner product, collapsing the distributional geometry of the embeddings. By contrast, the Wasserstein distance in Eq. (7) measures the minimum cost of transporting one subject's point cloud onto another's, preserving spatial arrangement in $\mathbb{R}^{N(H+1)}$. The optimal transport plan $P^*$ has a natural neuroimaging reading: entry $P^*_{ij}$ quantifies the fraction of mass transported from region $r_i$ of subject $s$ to region $r_j$ of subject $s'$, providing a soft geometry-aware alignment between connectomes.

*Embedding shift as a coupling proxy..* The quantity

$$\delta_s(r_i) = \|a_s^H(r_i) - a_s^0(r_i)\| \tag{12}$$

quantifies how strongly the functional profile of region $r_i$ is reshaped by structural propagation. Because the propagation operator (Eq. (6)) repeatedly averages each region's profile with its structural neighbours, $\delta_s(r_i)$ is small precisely when a region's function already resembles what its structural neighbourhood would predict, i.e. when structure and function are well coupled locally, and large when the region's function departs from that neighbourhood consensus, i.e. under local structure-function decoupling. $\delta_s$ is therefore a coupling proxy specific to this propagation-based embedding, not a drop-in substitute for other structure-function coupling measures in the literature (e.g. direct structural-functional connectivity correlation or communicability-based coupling), and should be read as such. Regions and subjects with large $\delta_s$ are natural candidates for investigating structure-function decoupling in clinical populations such as Alzheimer's disease, where white-matter integrity is selectively compromised. Section 3.4 tests this proxy directly by correlating $\bar{\delta}_s$ with subject age, not used to construct the kernel, under both the local and the fractional propagation operator.

*Lambda calibration..* The bandwidth $\lambda$ scales Wasserstein distances into the unit interval. We calibrate it on the training fold of each cross-validation split as $\lambda = 1/\bar{d}_{\text{train}}$, where $\bar{d}_{\text{train}}$ is the mean pairwise $W_1$ distance over training subjects, refined by a joint grid search over $(\lambda, C)$ around this reference value (Section 3.2). Crucially, both $\lambda$ and $C$ are re-estimated inside each outer fold so that no test-set information leaks into the kernel.

*Computational cost..* Exact $W_1$ has complexity $O(N^3 \log N)$ per pair; on a 100-region parcellation each pair requires approximately 1.6 ms. For a cohort of $S = 900$ subjects, computing the full distance matrix requires $\binom{900}{2} = 404{,}550$ pairwise evaluations ($\approx$11 min on a single core, well under a minute on a multi-core workstation utilizing parallelization). For exceptionally large cohorts or sub-millimetric parcellations, Sinkhorn regularisation [5] reduces this to near-linear complexity at the cost of an additional hyperparameter.

## 3. Experiments

### *3.1. A CN/MCI/AD cohort with a network-segregation-loss signal*

*Rationale..* The synthetic cohort is designed to test the sensitivity of the proposed graph representation to a structured, spatially distributed functional signal in a multi-class setting. We consider three ordered groups, Cognitively Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's Disease (AD), and generate the functional connectivity of each subject from an age-dependent model in which the group-related effect is expressed as a progressive loss of resting-state network segregation. In particular, within-network synchrony, task-positive/task-negative anticorrelation, and the contribution of polysynaptic structural paths are progressively reduced as a continuous function of subject age. This construction deliberately avoids a purely global rescaling of functional connectivity. The group signal therefore modifies the organisation of connectivity across functional systems rather than only its overall magnitude, producing a structured perturbation of the functional-connectivity matrix. The resulting benchmark provides a controlled setting in which the discriminative information is distributed across multiple regions and network-level relationships, while retaining a known continuous parameter governing signal strength. The three-group formulation also allows us to distinguish the information contained in the functional attributes from the contribution of structural propagation. In particular, because the same functional signal can be analysed both before and after graph propagation,

the benchmark provides a direct test of whether incorporating the structural topology changes the information captured by the Wasserstein representation. This separation is essential for assessing the proposed fractional propagation operator independently of the intrinsic discriminative power of the underlying functional-connectivity attributes.

*Generative model..* We generate a cohort of $S = 900$ subjects from six independently seeded structural templates. Each template is defined on the Schaefer-100 parcellation using the seven Yeo resting-state networks as communities [17], with within-network and between-network edge probabilities of 0.70 and 0.12, respectively. For each template, subject-specific structural connectivity matrices are generated by applying log-normal edge scaling, additive edge-level Gaussian noise with $\sigma_{\mathrm{SC}} = 0.08$, and random edge dropout. Each seed contributes 150 subjects. Subject ages are sampled independently from $\mathcal{U}(55, 90)$. The three clinical groups are assigned according to age quantiles after adding zero-mean Gaussian noise with standard deviation 8 years, with target proportions of 0.34, 0.33, and 0.33 for CN, MCI, and AD, respectively. The resulting cohort contains 306 CN, 300 MCI, and 294 AD subjects.

For each subject, the functional connectivity matrix is generated from a network-level block model. Let $B$ denote the baseline functional-connectivity matrix, whose entries are assigned according to the corresponding Yeo-network pair. Within-network pairs have mean connectivity 0.50; Dorsal Attention-Frontoparietal pairs have mean connectivity 0.20; Default-Default pairs have mean connectivity 0.15; task-positive-task-negative pairs have mean connectivity $-0.10$; and all remaining between-network pairs have baseline value 0.05. To incorporate subject-specific structural information, we add a polysynaptic contribution derived from the subject's pre-sparsification structural matrix $W$. The resulting functional-connectivity matrix is therefore given by a network-level component combined with a subject-specific structural contribution.

The strength of the network-level and structural signal is controlled by a continuous age-dependent factor. Let

$$\mathrm{age}_{\mathrm{frac}} = \frac{\mathrm{age} - 55}{90 - 55} \in [0, 1],$$

and define

$$\mathrm{signal}_{\mathrm{scale}} = \max(0.05,\ 1 - s\,\mathrm{age}_{\mathrm{frac}})\,, \tag{13}$$

where $s$ is the age-signal slope. The structured component of the functional-connectivity matrix is linearly contracted towards the unstructured baseline value 0.05 according to

$$C_{\mathrm{FC}} = 0.05 + \mathrm{signal}_{\mathrm{scale}} \left[ B + C_{\mathrm{struct}} - 0.05 \right]. \tag{14}$$

Finally, independent edge-level Gaussian noise with standard deviation $\sigma_{\mathrm{FC}} = 0.08$ is added to obtain the subject-specific functional-connectivity matrix.

We consider two values of the age-signal slope, $s = 0.6$ and $s = 0.3$, corresponding respectively to the full-strength and reduced-signal conditions. The latter is used as the primary classification condition, providing a more challenging setting in which the network-level structure of the signal is progressively weakened with age.

Clinical variables are generated independently of the structural and functional connectivity matrices from group-conditioned distributions. These variables include sex, MMSE, CSF $A\beta_{42}$, CSF pTau, and APOE4, and are used only for the downstream interpretability analysis.

Remark 3 shows that the fractional propagation operator can only introduce non-local interactions within the connected components of a subject's structural graph. We therefore explicitly assessed whether the fixed percentile threshold introduced in Section 2.2 preserves graph connectivity. At the default 70th percentile, the structural graph was fragmented into multiple components for every subject in a 100-subject check, with a mean of 5.9 components and a maximum of 8; none of the graphs was fully connected. We therefore lowered the threshold to the 50th percentile, which restored single-component connectivity for every subject checked while retaining a sparsified structural graph. All experiments reported below use this 50th-percentile threshold.

### *3.2. Evaluation protocol*

We use a nested stratified cross-validation (5-fold outer, 3-fold inner, joint $(\lambda, C)$ grid search per outer fold, Section 2.5), extended with the local-vs.-fractional propagation choice as an outer model-selection step: for the local operator (Eq. (6)) we select $H \in \{1, 2, 3\}$ by nested-CV balanced accuracy; for the fractional operator (Eq. (11)) we jointly select $H \in \{1, 2, 3\}$, $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and $\beta \in \{1, 2\}$ over the same grid. We compare against a raw-feature SVM on the concatenated upper triangles of SC and FC (Flat-SVM, no propagation) and the histogram-based WL-subtree kernel [18]

Table 1: Balanced accuracy (mean ± std over 5 outer folds), CN/MCI/AD cohort, $S = 900$, connectivity-validated 50th-percentile sparsification (Section 3.1). **Hard**: age-signal slope $s = 0.3$. **Full-strength**: $s = 0.6$, reported for comparison. WWL/fWWL: local/fractional propagation, $H$ (and $\alpha, \beta$) selected by nested CV (Section 3.2). FC-only/SC-only: propagation operator, resp. functional input, set to the identity - isolate whether the signal is already present in FC or SC alone. GraphTheory/ShortestPath: complete, independently-motivated external pipelines, not ablations of WWL. Chance level = 0.333.

| Method | Hard ($s = 0.3$) | Full-strength ($s = 0.6$) |
|---|---|---|
| WWL | $0.651 \pm 0.028$ | $0.657 \pm 0.023$ |
| fWWL | $0.654 \pm 0.028$ | $0.665 \pm 0.025$ |
| FC-only | $0.653 \pm 0.029$ | $0.658 \pm 0.027$ |
| ShortestPath | $0.641 \pm 0.031$ | $0.670 \pm 0.041$ |
| GraphTheory | $0.594 \pm 0.030$ | $0.619 \pm 0.034$ |
| WL-subtree | $0.480 \pm 0.041$ | $0.619 \pm 0.039$ |
| Flat-SVM | $0.467 \pm 0.041$ | $0.601 \pm 0.029$ |
| SC-only | $0.316 \pm 0.023$ | $0.328 \pm 0.015$ |

evaluated at the same nested-CV-selected $H$ as the local WWL operator, so that the four methods are compared on equal footing. Statistical significance is assessed with a permutation test ($n_{\text{perm}} = 1000$).

### *3.3. Classification results*

Table 1 and Figure 2 add four comparisons to the local/fractional pair. FC-only and SC-only isolate the functional and structural channels, respectively, by setting the propagation operator or the functional input to the identity. GraphTheory and ShortestPath provide complete, independently motivated alternative pipelines rather than ablations of WWL. SC-only remains close to chance in both conditions (0.316-0.328), confirming that the structural templates alone do not encode the group signal, as intended in the generative model. FC-only, in contrast, closely matches WWL and fWWL (0.653-0.658 versus 0.651-0.665): when the propagation operator is replaced by the identity, eliminating structural propagation entirely, the Wasserstein comparison of the raw functional connectivity profiles retains essentially the same classification performance. This indicates that, for the present benchmark, the discriminative information is already contained in the functional attributes and that structural propagation does not provide a measurable additional contribution. ShortestPath, based on an RBF-SVM applied to FC-

derived shortest-path-length histograms [4], gives comparable performance (0.641-0.670), while GraphTheory, based on a linear SVM applied to handcrafted network summary statistics, gives lower balanced accuracy (0.594-0.619). These comparisons further show that the observed performance is not specific to the inclusion of structural propagation, but is closely tied to how the functional connectivity profiles are represented and compared.

The results do not indicate a measurable benefit of structural propagation: WWL, fWWL, and FC-only are statistically indistinguishable under both signal conditions. Thus, for the present benchmark, the functional attributes already contain sufficient information for group discrimination, and incorporating the structural propagation operator does not produce a detectable change in classification performance (Section 3.4 examines this result in the context of the fractional extension).

A more pronounced difference emerges in the robustness to signal weakening. Methods based on distributional representations of functional connectivity, the Wasserstein point clouds of WWL/fWWL/FC-only and the histogram representation combined with a nonlinear kernel in ShortestPath, remain in the range 0.59-0.67 across the two signal conditions. By contrast, the methods that represent the propagated connectivity through either a flat feature vector (Flat-SVM) or a linear inner product of propagated histograms (WL-subtree) show a substantially larger performance reduction when the signal is weakened: balanced accuracy decreases from 0.601 to 0.467 for Flat-SVM and from 0.619 to 0.480 for WL-subtree, compared with a change from 0.657 to 0.651 for WWL. Within this controlled benchmark, these results therefore point to the representation and comparison of functional connectivity distributions, rather than structural propagation itself, as the main source of robustness to reduced signal strength.

Figure 3 visualises the two classification conditions in kernel-PCA space. The two-dimensional unsupervised projection shows substantial overlap between the three groups in both conditions, despite balanced accuracies of 0.651-0.665 obtained by the full-dimensional kernel classifier (Table 1). The figure should therefore be interpreted as a qualitative visualisation of the geometry induced by the kernel, rather than as a direct assessment of classification performance. The limited separation is also reflected in the kernel spectrum. The first three kernel-PCA components jointly account for only 0.34% of the centred kernel's total positive eigenvalue mass in both conditions, with approximately 0.11% contributed by each of PC1-PC3. Thus, the two- and three-dimensional projections shown in Figures 3 and 4 retain

Table 2: Confusion-matrix summary. Extreme confusion (Extr. conf.): mean of $P(\hat{y}=\text{CN} \mid y=\text{AD})$ and $P(\hat{y}=\text{AD} \mid y=\text{CN})$. MCI recall: $P(\hat{y}=\text{MCI} \mid y=\text{MCI})$.

| | Hard | | Full-strength | |
|---|---|---|---|---|
| Method | Extr. conf. | MCI recall | Extr. conf. | MCI recall |
| WWL | 5% | 42% | 4% | 45% |
| fWWL | 4% | 43% | 4% | 46% |
| FC-only | 4% | 43% | 4% | 44% |
| ShortestPath | 4% | 42% | 3% | 52% |
| GraphTheory | 6% | 40% | 5% | 42% |
| WL-subtree | 14% | 35% | 8% | 42% |
| Flat-SVM | 12% | 36% | 6% | 38% |
| SC-only | 26% | 41% | 35% | 23% |

only a small fraction of the variation represented by the full $S \times S$ kernel matrix. The discrepancy between the weak low-dimensional separation and the substantially higher classification accuracy indicates that the discriminative information is not concentrated in the leading kernel-PCA directions. Rather, classification exploits information distributed across a larger portion of the kernel representation. This is consistent with the relatively diffuse similarity structure produced by the kernel parameterisation described in Section 2.5.

*Methods comparison..* Balanced accuracy does not characterise the types of classification errors made by a method. Table 2 therefore summarises the pooled out-of-fold confusion matrices for the hard condition using two derived quantities: the rate of confusion between the two extreme groups (AD predicted as CN or CN predicted as AD) and MCI recall. The extreme-confusion rate is low for WWL, fWWL, FC-only, ShortestPath, and GraphTheory (4-6%), whereas it is higher for WL-subtree and Flat-SVM (14% and 12%, respectively). Thus, although their aggregate balanced accuracies remain above chance, these two methods show a different error profile, with more frequent confusion between the two extreme classes. SC-only exhibits a distinct behaviour: its extreme-confusion rate is 26%, while MCI is predicted for 41% of samples, indicating a strongly non-uniform prediction pattern despite its near-chance balanced accuracy. The same quantities for the full-strength condition are reported in Table 2.

The qualitative differences between methods remain, although their mag-

nitude changes. In particular, WWL, fWWL, FC-only, ShortestPath, and GraphTheory retain low extreme-confusion rates (3-5%), whereas WL-subtree and Flat-SVM remain higher. ShortestPath has the highest MCI recall in the full-strength condition (52%), compared with 45% for WWL, consistent with its higher aggregate balanced accuracy in this condition (0.670 versus 0.657; Table 1). This does not establish a systematic difference between the two methods: across the five paired outer folds, WWL has lower balanced accuracy in four folds, with a mean difference of 1.2 percentage points, but neither a paired $t$-test ($p = 0.27$) nor a Wilcoxon signed-rank test ($p = 0.31$) detects a significant difference. Given the five-fold sample size, these tests have limited power, so the comparison is best regarded as inconclusive.

The main difference between the two methods concerns stability across signal strength. WWL changes by only 0.6 percentage points between the hard and full-strength conditions ($0.651 \rightarrow 0.657$), whereas ShortestPath changes by 2.9 percentage points ($0.641 \rightarrow 0.670$). Thus, within this benchmark, the results do not establish a classification advantage of WWL over ShortestPath at either operating point; rather, they indicate that the WWL performance is less sensitive to the reduction in signal strength.

### *3.4. Regularized fractional propagation: a signal-dependent effect*

To assess whether the effect of non-local propagation depends on the spatial structure of the underlying signal, we extended the benchmark beyond the CN/MCI/AD block cohort described in Section 3.1. Two additional synthetic cohorts were generated using the same structural-connectivity framework and sample size ($S = 900$), but with different functional signal structures. The first, referred to as the *communicability* cohort, introduces a diffuse whole-brain signal through a uniform age-dependent modulation of structural-functional coupling. The second retains the block-structured functional organisation of the CN/MCI/AD cohort but uses a reduced age-signal slope. These additional cohorts provide complementary settings in which to test whether the non-local reach of the fractional operator is sensitive to the spatial scale of the signal. All three cohorts use the connectivity-validated 50th-percentile structural sparsification.

Table 3 reports exhaustive sweeps over the $(H, \alpha, \beta)$ grid for the three synthetic benchmarks. For the two block-based cohorts, the full 30-point grid produces only small variations in balanced accuracy (less than 1.5 percentage points), with no consistent dependence on either $\alpha$ or $\beta$ (Figure 5). The best fractional configurations therefore provide only a small change relative to

Table 3: Fractional vs. local propagation, full $(H, \alpha, \beta)$ grid ($H \in \{1, 2, 3\}$, $\alpha \in \{0.1, \ldots, 0.9\}$, $\beta \in \{1, 2\}$, 30 combinations per cohort), best-of-grid reported. All three cohorts at $S = 900$, 50th-percentile sparsification.

| Cohort | Local (best $H$) | Fractional (best grid point) | $\Delta$ |
|---|---|---|---|
| Communicability (uniform signal) | $0.500 \pm 0.036$ | $0.498 \pm 0.040$ | $-0.002$ |
| Block, full-strength | $0.657 \pm 0.023$ | $0.665 \pm 0.025$ | $+0.008$ |
| Block, hard | $0.651 \pm 0.028$ | $0.654 \pm 0.028$ | $+0.003$ |

Table 4: Pearson correlation of whole-brain mean embedding shift with subject age, local vs. fractional propagation ($H = 3$, $\alpha = 0.7$).

| Cohort | $r$ (local) | $r$ (fractional) | $\Delta r$ |
|---|---|---|---|
| Communicability (uniform signal) | 0.500 | 0.512 | $+0.012$ |
| Block, full-strength | 0.967 | 0.956 | $-0.011$ |
| Block, hard | 0.880 | 0.797 | $-0.083$ |

the best local configuration: $\Delta = +0.008$ for the full-strength condition and $\Delta = +0.003$ for the hard condition. On the communicability benchmark, the fractional configurations instead show a monotonic increase in accuracy as $\alpha \to 1$, corresponding to the limiting regime in which the fractional operator approaches the local propagation operator. Thus, across the three benchmarks considered, we find no evidence that non-local structural propagation provides a systematic advantage for discrete classification.

The comparison changes when the propagated representation is evaluated through a continuous readout. We correlate the whole-brain mean embedding shift $\bar{\delta}_s = \frac{1}{N} \sum_i \delta_s(r_i)$ with each subject's age, which is not used in kernel construction. For consistency across cohorts, this analysis uses $H = 3$ and compares the local operator with the fractional operator at $\alpha = 0.7$, $\beta = 1$ (Table 4).

On the communicability benchmark, the fractional operator produces a small but consistent increase in the age correlation ($r = 0.500$ to $0.512$). A paired bootstrap over subjects (2000 resamples) gives a mean difference of $\Delta r = 0.012$ with a 95% confidence interval of $[0.005, 0.020]$. The same direction is observed in nested-CV linear-SVR prediction using the 100-dimensional per-region shift vector, with $R^2_{\mathrm{CV}} = 0.163$ for local propagation and 0.180-0.182 for fractional propagation at $\alpha \in \{0.5, 0.7\}$. In contrast, the fractional operator reduces the age correlation for both block cohorts, from

Table 5: Biomarker associations with embedding shift, hard cohort. $r$: aggregate Pearson correlation of whole-brain mean shift with the biomarker (not used to build the kernel). $R^2_{\mathrm{CV}}$, $r_{\mathrm{CV}}$: nested-CV linear-SVR on the 100-dimensional per-region shift vector. Local: $H = 3$. Fractional: $H = 3, \alpha = 0.7, \beta = 1$.

| | Local | | | Fractional | | |
|---|---|---|---|---|---|---|
| Biomarker | $r$ | $R^2_{\mathrm{CV}}$ | $r_{\mathrm{CV}}$ | $r$ | $R^2_{\mathrm{CV}}$ | $r_{\mathrm{CV}}$ |
| Age | 0.880 | 0.915 | 0.957 | 0.797 | 0.861 | 0.928 |
| CSF A$\beta_{42}$ | −0.572 | 0.357 | 0.598 | −0.528 | 0.337 | 0.581 |
| CSF pTau | 0.548 | 0.332 | 0.579 | 0.503 | 0.308 | 0.559 |
| MMSE | −0.471 | 0.236 | 0.500 | −0.428 | 0.225 | 0.488 |

0.967 to 0.956 in the full-strength condition and from 0.880 to 0.797 in the hard condition.

These results indicate that the effect of non-local propagation depends on the spatial scale of the injected signal. When the functional perturbation is diffuse across the whole brain, fractional propagation can preserve or slightly enhance the correspondence between the propagated representation and the continuous signal. When the perturbation is structured around specific resting-state networks, extending propagation to distant structural neighbours can instead reduce this correspondence. This interpretation is consistent with a scale-matching view of the fractional operator: its power-law propagation can be beneficial when the spatial extent of the operator is compatible with that of the underlying signal, but can dilute more spatially localised structure. This mechanism remains a hypothesis suggested by the controlled benchmarks rather than a demonstrated general property of the method.

### *3.5. Biomarker associations with embedding shift*

Table 5 extends the age-only comparison of Table 4 to three additional biomarkers available in the cohort: MMSE, CSF A$\beta_{42}$, and CSF pTau. APOE4 status is also available in the simulated metadata but is not considered here. The analysis is performed on the hard condition using local ($H = 3$) and fractional ($H = 3, \alpha = 0.7, \beta = 1$) propagation, with the same aggregate-correlation and nested-CV linear-SVR procedures used for the age analysis.

The embedding shift shows the same ordering of association strength across the four variables under both propagation schemes (Figure 6): age has

the strongest association, followed by CSF A$\beta_{42}$ and CSF pTau, with MMSE showing the weakest association. For local propagation, the corresponding aggregate correlations are $|r| = 0.880$, 0.572, 0.548, and 0.471, respectively. The fractional operator reduces the magnitude of all four associations, with the largest relative change observed for age. The nested-CV results show the same pattern: fractional propagation yields lower $R^2_{\mathrm{CV}}$ and $r_{\mathrm{CV}}$ for every variable. Thus, the effect observed for age in Table 4 extends consistently to the other simulated biomarkers rather than being specific to a single readout.

More directly, these results support the signal-scale interpretation developed in Section 3.4. All four variables are generated as downstream correlates of the same age-dependent functional signal, whose spatial structure is defined by the resting-state-network block model (Section 3.1). The fact that fractional propagation reduces the correspondence with all four readouts is therefore consistent with the interpretation that non-local averaging dilutes a signal whose discriminative structure is concentrated within specific functional networks. This should be regarded as a property of the present generative model rather than as evidence for a general disadvantage of fractional propagation.

*Ordinal consistency across the three diagnostic groups..* Because CN, MCI and AD are ordered by construction, we additionally examined whether the regional embedding shift follows the same ordering across adjacent diagnostic groups. For each region $r_i$, we computed the difference in mean shift between groups,

$$\Delta\delta(r_i) = \bar{\delta}_{g_1}(r_i) - \bar{\delta}_{g_0}(r_i),$$

for the CN→MCI, MCI→AD, and CN→AD contrasts.

Figure 7 shows positive shifts for all three contrasts under local propagation. The mean regional differences are 0.164 for CN→MCI, 0.148 for MCI→AD, and 0.312 for CN→AD. The latter is close to the sum of the two adjacent-group differences, consistent with an approximately additive increase in embedding shift across the three ordered groups. All 100 regions are individually significant after Benjamini-Hochberg correction ($q < 0.05$) in each contrast.

Figure 8 extends this analysis from individual regions to resting-state networks and compares local and fractional propagation. The CN < MCI < AD ordering is preserved in every network under both operators. The corresponding regional contrasts under fractional propagation are also positive and FDR-significant across all regions, with mean differences of 0.149, 0.133,

and 0.282 for CN→MCI, MCI→AD, and CN→AD, respectively. These values are approximately 10% smaller than their local counterparts, consistent with the lower biomarker associations reported in Table 5. Thus, the local and fractional operators recover the same ordinal and regional structure, with the main difference being the magnitude of the embedding shift.

The ranking of the network-level shifts is also preserved between propagation schemes, with Limbic and Salience showing the largest absolute shifts and Default and Frontoparietal the smallest. This ranking should not be interpreted as a measure of clinical relevance: the functional signal in the present benchmark is defined by a spatially distributed reduction in within-network segregation rather than by a network-specific pathological effect. Consequently, the magnitude of the embedding shift reflects the construction of the synthetic signal and the resulting representation, not the known clinical importance of individual resting-state networks.

Finally, the strong age association under local propagation ($R^2_{\mathrm{CV}} = 0.915$) reflects the close, deterministic dependence of the simulated signal-scale factor on age at the hard slope (Eq. (13)), not an independent structure-function coupling effect. This strong relationship is therefore a property of the generative model and should not be interpreted as an estimate of the strength of an analogous age-embedding relationship in empirical data.

### *3.6. Limitations*

The distinction between the hard and full-strength conditions corresponds to two age-signal slopes ($s = 0.3$ and $s = 0.6$), rather than to a continuous gradient of signal difficulty. The permutation test was performed only for the primary hard, local-propagation configuration and was not repeated for every configuration reported in Tables 1 and 3. The biomarker analysis in Table 5 is likewise restricted to the hard condition and to the four continuous readouts considered here; APOE4 status is available in the simulated metadata but was not analysed. The regional and network-level analyses reported above provide additional checks of the embedding behaviour in the three-group setting, but further analyses of regional specificity and negative controls could be considered in future work. Most importantly, all results are based on controlled synthetic cohorts. Validation on independent empirical neuroimaging data is therefore required to assess whether the observed behaviour of the Wasserstein representation and the fractional propagation operator extends beyond the present generative models.

## 4. Conclusion

We have presented a connectome Wasserstein Weisfeiler-Lehman (WWL) kernel extended with a regularized fractional-Laplacian propagation operator following [16]. The fractional construction allows non-local structural propagation within each connected component, with edge influence decaying with structural distance rather than being truncated at a prescribed number of propagation steps. At the same time, the construction remains anchored to the subject-specific structural graph, as established by Eq. (10) and Remark 3. We evaluated the framework on a three-group synthetic cohort (CN, MCI, and AD) in which the functional attributes encode a graded, network-structured loss of resting-state segregation. In this setting, the main contribution to classification performance comes from the Wasserstein comparison of the functional attribute distributions rather than from structural propagation itself. The identity-propagation control gives performance comparable to both local and fractional WWL, whereas the structure-only control remains close to chance. Thus, for the present benchmark, the structural propagation step does not provide a measurable additional contribution to classification. This result is important for interpreting the fractional extension: the absence of a classification gain cannot be attributed simply to an inappropriate choice of fractional parameters, since the underlying functional signal is already accessible without structural propagation. The comparison with alternative representations leads to a similarly specific conclusion. WWL, fractional WWL, FC-only, and ShortestPath achieve comparable balanced accuracies across the tested conditions, while the performance of the latter methods varies more with the strength of the injected functional signal. In particular, WWL changes by only 0.6 percentage points between the hard and full-strength conditions, compared with a 2.9 percentage-point change for ShortestPath. The paired comparison across the five outer folds does not establish a statistically significant difference between the two methods. These results therefore do not support a claim of superior classification accuracy for WWL; rather, within this controlled benchmark, they indicate that the Wasserstein representation provides competitive performance with comparatively limited sensitivity to the strength of the functional signal. The regularized fractional operator was further evaluated over its full $(H, \alpha, \beta)$ parameter grid on three synthetic benchmarks with different spatial structures of the functional signal. No systematic classification improvement over the best local configuration was observed. A different behaviour emerged when

the propagated representation was evaluated through continuous readouts. For the diffuse communicability benchmark, fractional propagation slightly increased the correlation between the embedding shift and age, whereas for the block-structured benchmarks it reduced the corresponding correlation. The effect was therefore dependent on the spatial scale of the injected signal rather than on its overall strength alone. This provides a concrete and testable interpretation of the fractional extension: non-local propagation may be useful when its spatial reach is compatible with the scale of the underlying signal, whereas extending propagation beyond that scale can dilute more localized structure. The present experiments do not establish this principle as a general property, but provide a controlled setting in which it can be tested.

An additional methodological observation concerns the structural sparsification used to define the propagation graph. Because fractional propagation cannot cross disconnected structural components, the connectivity of the sparsified graph directly constrains the effective spatial reach of the operator. The connectivity analysis showed that the original 70th-percentile threshold fragmented the structural graphs, whereas the 50th-percentile threshold produced a single connected component for all subjects in the validation sample. All experiments reported here therefore use the connectivity-validated threshold. This check is relevant independently of the classification results, since it separates the intended behaviour of the fractional operator from an artefact introduced by graph construction. Beyond classification, the embedding shift provides a direct subject- and region-level quantity for interrogating the geometry induced by the kernel. In the three-group benchmark, the regional shift increases consistently across CN, MCI, and AD, under the coupling interpretation of Eq. (12), a progressive weakening of local structure-function coupling, with the same ordinal ordering observed across all seven resting-state networks under both local and fractional propagation. The shift also shows associations with age, CSF A$\beta_{42}$, CSF pTau, and MMSE in the hard condition. These associations are expected to some extent from the construction of the synthetic cohort, where the functional signal and the group-conditioned metadata are linked through the generative model. They should therefore be interpreted as demonstrations of the readout available from the representation, rather than as evidence of clinical validity. Importantly, the same representation used for classification produces these spatially resolved quantities without introducing a separate predictive model for each readout. Taken together, the results support a limited but well-defined conclusion.

The WWL framework provides a distributional representation of continuous functional attributes on a subject-specific structural graph, while its embedding shift offers an interpretable way to examine how those attributes change across subjects and regions. On the synthetic benchmarks considered here, classification performance is primarily determined by the functional attributes and their Wasserstein comparison, whereas structural propagation has a smaller and signal-dependent role. The fractional extension consequently should not be regarded as a universally beneficial replacement for local propagation; its potential value appears to depend on the spatial scale at which the underlying signal is organised. Several limitations define the scope of these conclusions. The hard and full-strength conditions correspond to only two values of the age-signal slope, rather than to a continuous difficulty gradient. Permutation testing was performed for the primary hard, local-propagation configuration and was not repeated for every configuration in the experimental grid. The biomarker analysis was restricted to the hard condition and to the four continuous readouts considered above. Most importantly, all experiments were performed on controlled synthetic cohorts, whose functional and clinical variables are generated according to prescribed statistical relationships. Validation on independent empirical neuroimaging data is therefore necessary to determine whether the observed behaviour of the Wasserstein representation and the fractional propagation operator extends beyond these generative models. Future work should also test the signal-locality interpretation more directly by constructing functional perturbations with controlled spatial support, for example by concentrating the network-segregation signal within selected resting-state networks. Such experiments would provide a direct test of whether the spatial scale of the injected signal predicts the relative behaviour of local and fractional propagation, and would help determine when the additional flexibility of the fractional operator is warranted.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data and code availability

An implementation of the connectome WWL kernel framework is available at `https://github.com/theMIDAgroup/WWL_connectome_kernel.git`.


## Acknowledgements

The authors acknowledge the financial support of the "Hub Life Science - Digital Health (LSH-DH) PNC-E3-2022-23683267 - Progetto DHEAL-COM - CUP: D33C22001980001", granted by the Italian Ministero della Salute within the framework of the Piano Nazionale Complementare to the "PNRR Ecosistema Innovativo della Salute - Codice univoco investimento: PNC-E.3". The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. The authors also acknowledge the support obtained by INdAM under the GNCS research project "Metodi kernel per l'integrazione della connettività cerebrale multimodale", Code CUP E53C25002010001.


## Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used Claude (Anthropic) to assist in drafting and revising portions of the manuscript text. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

## References


[1] Yu, Meichen and Sporns, Olaf and Saykin, Andrew J., et al. Multilayer structural-functional connectome changes are associated with amyloid-$\beta$ and tau pathologies in Alzheimer's Disease, Alzheimer's & Dementia, **20**,(S2), e091424, (2024)

[2] Battiston, F., Nicosia, V., Chavez, M., Latora, V.: Multilayer motif analysis of brain networks. Chaos **27**(4), 047404 (2017)

[3] Boccaletti, S., Bianconi, G., Criado, R., et al.: The structure and dynamics of multilayer networks. Physics Reports **544**(1), 1-122 (2014)

[4] Borgwardt, K.M., Kriegel, H.P.: Shortest-path kernels on graphs. In: Proc. 5th IEEE Int. Conf. on Data Mining (ICDM), pp. 74-81 (2005)

[5] Cuturi, M.: Sinkhorn distances: lightspeed computation of optimal transport. In: Advances in NeurIPS 26, pp. 2292-2300 (2013)

[6] De Domenico, M., Solé-Ribalta, A., Cozzo, E., et al.: Mathematical formulation of multilayer networks. Phys. Rev. X **3**, 041022 (2013)

[7] Douglas, B.L.: The Weisfeiler-Lehman method and graph isomorphism testing. arXiv:1101.5211 (2011)

[8] Kivelä, M., Arenas, A., Barthelemy, M., et al.: Multilayer networks. Journal of Complex Networks **2**(3), 203-271 (2014)

[9] Breedt LC, et al. Structure-function multilayer network integration and cognition in multiple sclerosis. Netw Neurosci. 2026 Apr 22;10(2):400-417. doi: 10.1162/NETN.a.545. PMID: 42039096; PMCID: PMC13108497.

[10] Honey, C.J., et al.: Predicting human resting-state functional connectivity from structural connectivity. PNAS **106**(6), 2035-2040 (2009)

[11] Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: 5th ICLR (2017)

[12] Kriege, N.M., Johansson, F.D., Morris, C.: A survey on graph kernels. Applied Network Science **5**(1), 6 (2020)

[13] Li, Q., Han, Z., Wu, X.M.: Deeper insights into graph convolutional networks for semi-supervised classification. In: Proc. AAAI **32**(1) (2018)

[14] Nikolentzos, G., Siglidis, G., Vazirgiannis, M.: Graph kernels: a survey. JAIR **72**, 943-1027 (2021)

[15] Ong, C.S., Mary, X., Canu, S., Smola, A.J.: Learning with non-positive kernels. In: Proc. ICML (2004)

[16] Filippo, A., Mazza, M.: Spectral and computational aspects of a regularized fractional Laplacian for non-local diffusion on graphs. J. Numer. Math., aop (2026).

[17] Schaefer, Alexander and Kong, Ru and Gordon, Evan M and Laumann, Timothy O and Zuo, Xi-Nian and Holmes, Avram J and Eickhoff, Simon B and Yeo, B T Thomas, Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. Cerebral Cortex,(2018)

[18] Shervashidze, N., Schweitzer, P., van Leeuwen, E.J., Mehlhorn, K., Borgwardt, K.M.: Weisfeiler-Lehman graph kernels. JMLR **12**, 2539-2561 (2011)

[19] Suárez, L.E., Markello, R.D., Betzel, R.F., Misic, B.: Linking structure and function in macroscale brain networks. Trends Cogn. Sci. **24**(4), 302-315 (2020)

[20] Togninalli, M., Ghisu, E., Llinares-López, F., Rieck, B., Borgwardt, K.: Wasserstein Weisfeiler-Lehman graph kernels. In: Advances in NeurIPS 32. arXiv:1906.01277 (2019)

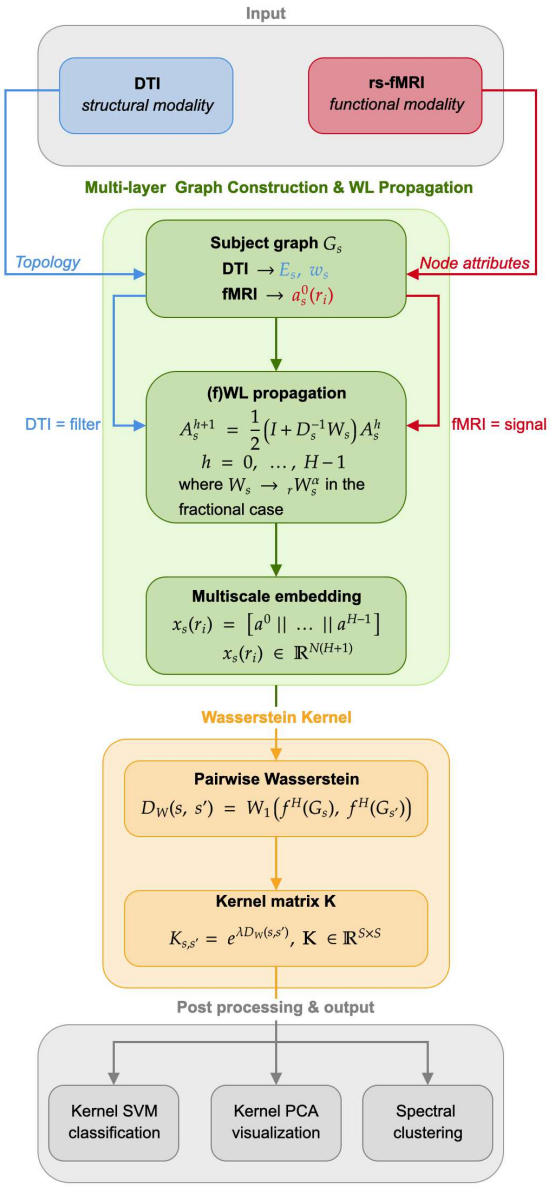
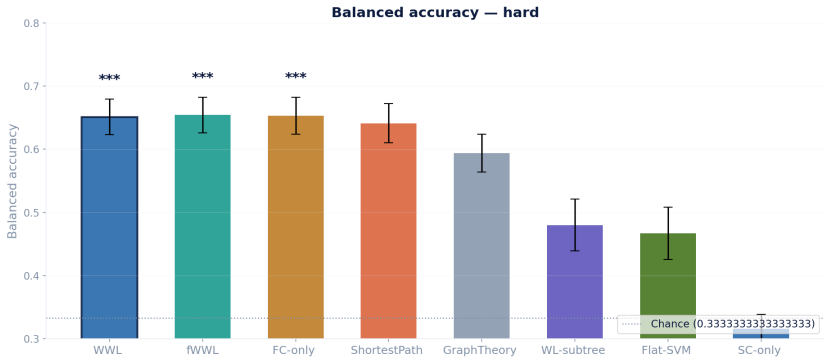
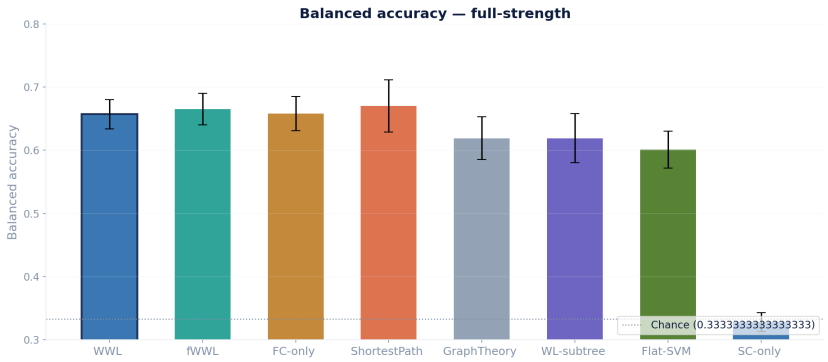



Figure 1: Overview of the proposed framework. *Top (green)*: multi-layer graph construction and WL propagation. DTI defines the propagation topology acting as a structural filter, while rs-fMRI provides the initial node attributes acting as the propagated signal. Iterative neighbourhood aggregation yields the multi-scale embedding representing each subject as a point cloud of $N$ vectors. *Middle (orange)*: pairwise Wasserstein distances are assembled into the symmetric kernel matrix. *Bottom (grey)*: the kernel is the direct input to any kernel-based downstream task.




(a) Hard ($s = 0.3$)




(b) Full-strength ($s = 0.6$)

Figure 2: Balanced accuracy, CN/MCI/AD cohort (Table 1). WWL, fWWL, FC-only and ShortestPath cluster together and stay comparatively robust as the signal weakens; Flat-SVM and WL-subtree collapse toward SC-only's chance level. Stars (hard condition only): permutation-test significance for the three Wasserstein-kernel methods (WWL, fWWL, FC-only); not run for the remaining methods.

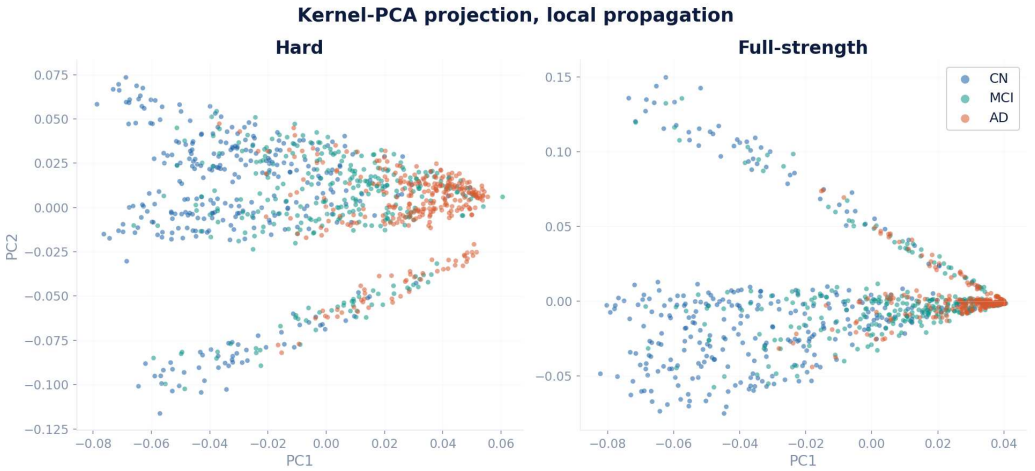



Figure 3: Unsupervised kernel-PCA projection of the WWL kernel (local propagation), hard vs. full-strength. Both conditions show substantial visual overlap between groups but not the same kind of structure: full-strength shows one tightly-converging CN→MCI→AD arm alongside a diffuse, poorly-separated cloud, while hard shows a more uniformly diffuse gradient throughout, without a comparably tight arm. Both panels also show a two-cluster ("two-arm") structure independent of diagnostic group, plausibly reflecting the six independent structural templates the cohort is pooled from (Section 3.1) rather than group membership.

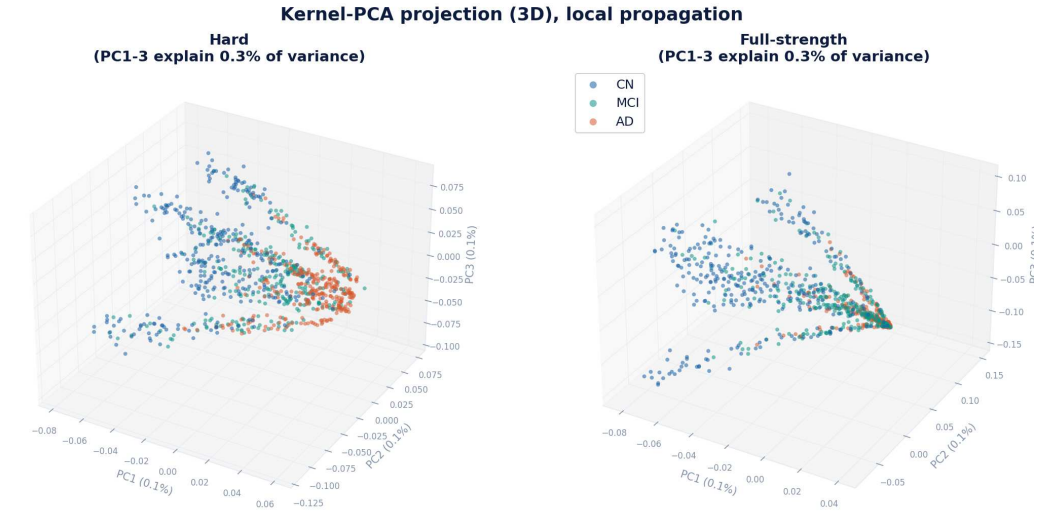



Figure 4: The same projection in 3D (PC1-PC3). The apparent "two-arm" split of Figure 3 is consistent with one continuous curved manifold seen edge-on in 2D, rather than two genuinely separate clusters.

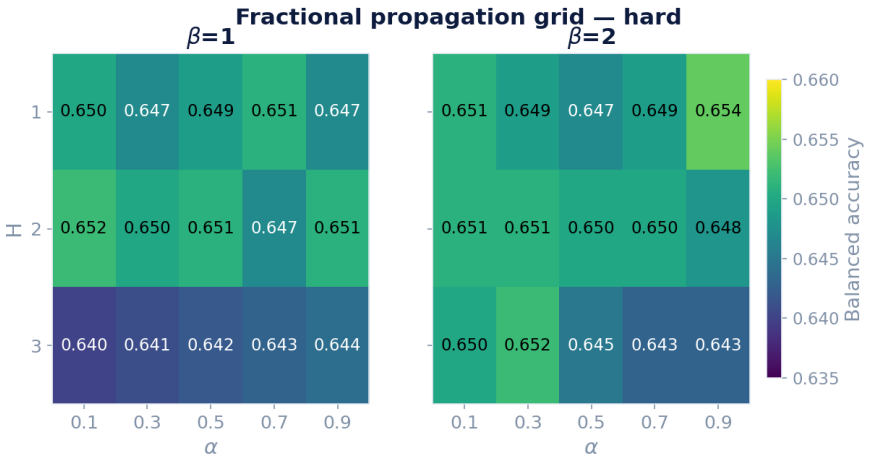



Figure 5: Balanced accuracy over the full $(H, \alpha, \beta)$ grid, hard cohort. No cell exceeds the local-propagation baseline ($H = 2$, 0.651) by more than one standard deviation, and no coherent trend in $\alpha$ is visible.

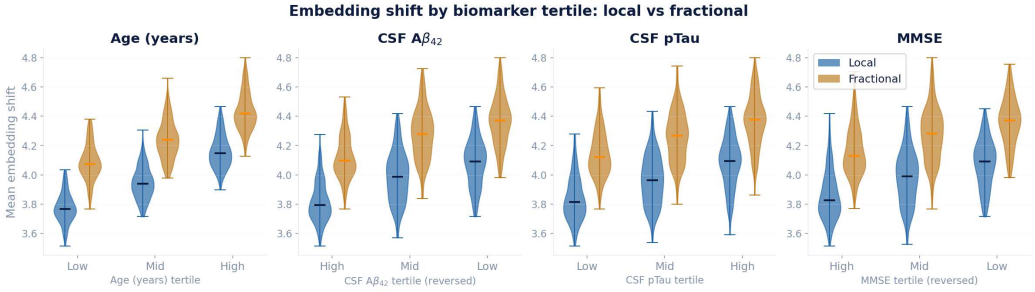



Figure 6: Mean embedding shift by biomarker tertile, local (blue) vs. fractional (orange) propagation, hard cohort. Tertiles for CSF A$\beta_{42}$ and MMSE are shown in reversed order (High-Mid-Low) so that, in every panel, left-to-right tracks increasing severity. Deliberately shown as tertile violins rather than a continuous regression fit, which would overstate the precision a synthetic-data correlation supports. Fractional shift sits systematically higher in absolute level but separates tertiles slightly less than local, consistent with Table 5's lower $R^2_{\text{CV}}$ for fractional despite the higher baseline.

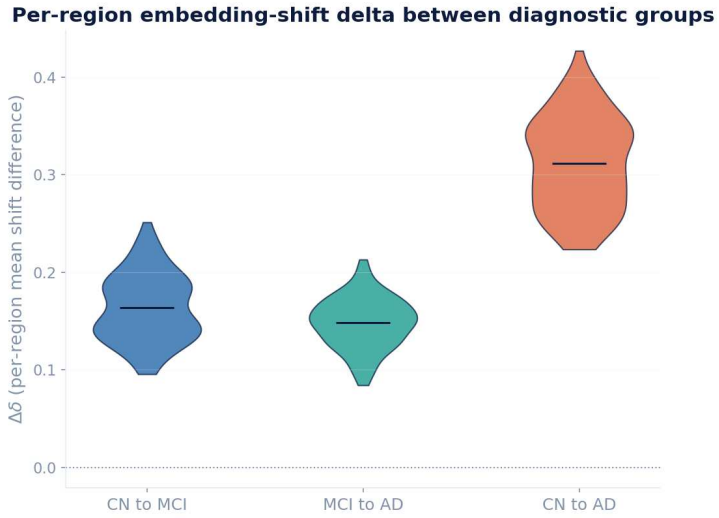



Figure 7: Per-region embedding-shift delta $\Delta\delta$ between diagnostic groups (local propagation, hard cohort; $N = 100$ regions per violin, every region FDR-significant, Benjamini-Hochberg $q < 0.05$, in all three contrasts).

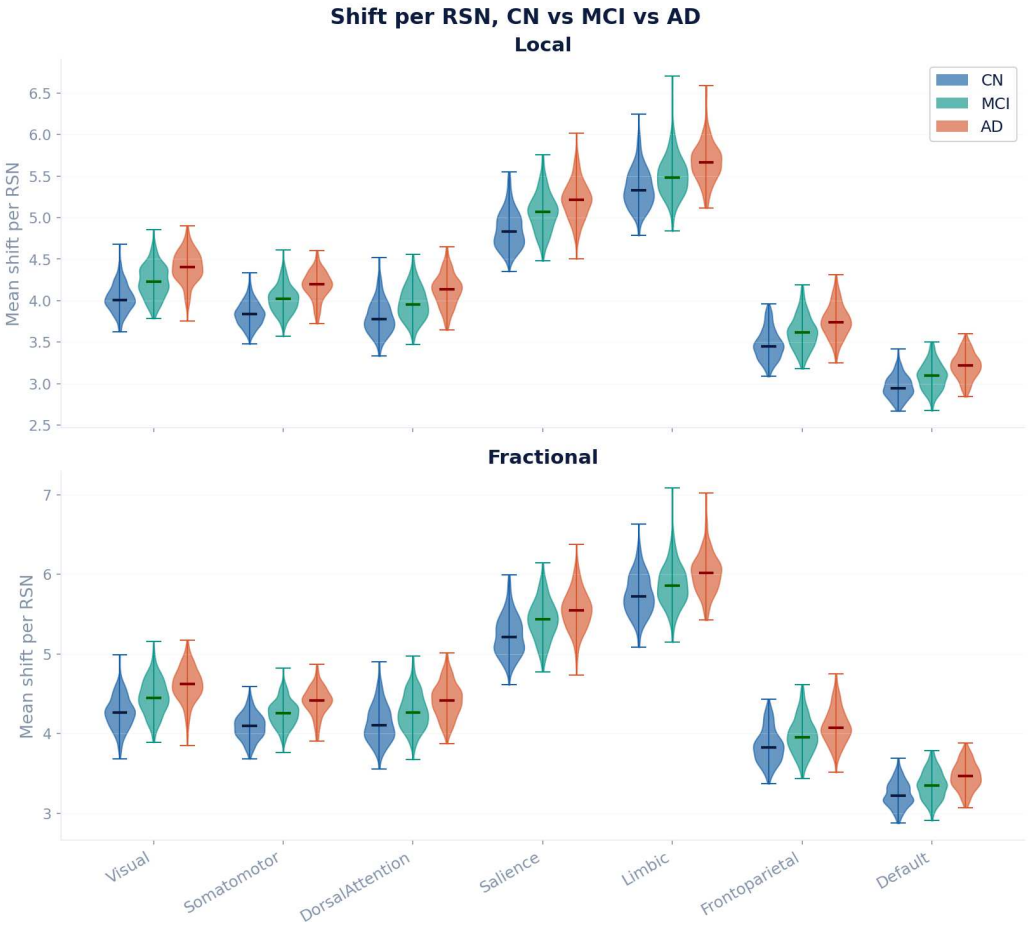



Figure 8: Per-subject mean shift by resting-state network, all three diagnostic groups, local (top) vs. fractional (bottom) propagation, hard cohort. Every network shows the same CN $<$ MCI $<$ AD ordering under *both* operators, with the same ranking of absolute magnitude (Limbic and Salience largest, Default and Frontoparietal smallest) - fractional shift is uniformly $\sim$10% smaller in this per-network breakdown, consistent with Table 5, but the qualitative pattern is unchanged.