
Using persistent homology to understand dimensionality reduction in resting-state fMRI

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 Evaluating the success of a manifold learning method remains a challenging problem, especially for methods adapted to a specific application domain. The present work investigates shared geometric structure across different dimensionality reduction (DR) algorithms within the scope of neuroimaging applications. We examine reduced-dimension embeddings produced by a representative assay of dimension reductions for brain data (“brain representations”) through the lens of persistent homology, making statistical claims about topological differences using a recent topological bootstrap method. We cluster these methods based on their induced topologies, finding feature type and number — rather than reduction algorithm — as the main drivers of observed topological differences.

11 1 Introduction

12 The present work investigates shared geometric structure across different dimensionality reduction algorithms within the scope of neuroimaging applications. For most applications, a “dimensionality reduction” is any of a large class of methods that make inferences about structures underlying some data, typically to represent this data in both a more efficient and more interpretable way. Many dimensionality reduction (DR) problems can be equivalently formulated as “manifold learning” problems (i.e., estimating the manifold from which a dataset was sampled), and we will use the terms synonymously. Efforts to understand theoretical and empirical relationships between DR methods remain active^{1–10}.

20 The difficulty of relating dimensionality reductions can be compounded within specific application domains because methodologies often branch into variably specialized use cases. Nonetheless, specific use cases can also suggest more stringent criteria by which to compare dimension reduction outcomes. In functional neuroimaging, specialized dimension reduction algorithms proliferate the field, bridging disparate use cases, design philosophies, and biological motivations^{11,12}. These DR algorithms share the goal of extracting networks of functional activity from resting-state fMRI brain data, and we will refer to them throughout as “brain representations.” We compare brain representations in terms of the topologies they induce on a single set of shared data (“subject space”) and the statistical robustness of the differences between them. We measure these topological statistics through persistent homology¹³ and the related topological bootstrap^{14,15}.

30 Problem Statement: Brain Representations & Subject Space

31 Our primary goal is to compare the structural changes in a single neuroimaging dataset under a variety of brain representations. Because the structure of subject space in the original (un-reduced) data is unknown and impractical to compute, it is not feasible to grade brain representations’ quality

34 by structure preservation. Instead, we group brain representations based on the similarity of the
 35 subject-space structures they induce in reduced data.

36 We frame brain representation as a manifold learning problem, and we frame comparisons between
 37 them as comparisons between estimated manifolds. We suppose our data lies on some manifold
 38 $\mathcal{S} \hookrightarrow \mathbb{R}^D$ ("subject space"), from which we extract a finite dataset \widehat{S} of N samples. In this
 39 work, \mathcal{S} consists of resting-state fMRI scans, and \widehat{S} is the Human Connectome Project (HCP)
 40 Young Adult dataset; we then have $D \sim 10^8$ and $N \sim 10^3$. A **brain representation** is any
 41 mapping $\widehat{\varphi}_i : \widehat{S} \rightarrow \mathbb{R}^{d_i}$ with $d_i \ll D$; for any brain representation, we define its corresponding
 42 **induced subject space** $\widehat{S}_i = \widehat{\varphi}_i(\widehat{S})$. We compare brain representations $\widehat{\varphi}_i$ and $\widehat{\varphi}_j$ by comparing the
 43 persistent homology of their induced subject spaces \widehat{S}_i and \widehat{S}_j . To link the persistent homology and
 44 manifold learning investigations, we make a key modeling assumption: there exists a local extension
 45 $\varphi_i : \mathcal{S} \rightarrow \mathbb{R}^{d_i}$ of $\widehat{\varphi}_i$ that is a submersion in some neighborhood of $\widehat{S} \subset \mathcal{S}$. This assumption requires
 46 that dimensionality reductions behave consistently on unseen data near training data, and constitutes
 47 only a mild smoothness assumption on $\widehat{\varphi}_i$. Under this modeling assumption, we may assert the
 48 existence of a manifold $\mathcal{S}_i \hookrightarrow \mathbb{R}^{d_i}$ containing \widehat{S}_i such that the following diagram commutes:

$$\begin{array}{ccc} \widehat{S} & \xrightarrow{\widehat{\varphi}_i} & \widehat{S}_i \\ \downarrow & & \downarrow \\ \mathcal{S} & \xrightarrow{\varphi_i} & \mathcal{S}_i \end{array}$$

49 While the smoothness assumption on $\widehat{\varphi}_i$ is easily met by most DR algorithms, the connection between
 50 the persistent homology of \widehat{S}_i and the manifold \mathcal{S}_i depends heavily on properties of the manifold
 51 sampling \widehat{S} .

52 To compare brain representations φ_i and φ_j , we compute dissimilarity metrics for all pairs of
 53 points in \widehat{S}_i and \widehat{S}_j and examine the resulting Vietoris-Rips complex in each space. This approach
 54 allows flexibility in the data and dissimilarities under consideration while still allowing claims about
 55 DR-induced topological differences.

56 **Related Work: Persistent Homology & Dimensionality Reduction**

57 Most comparisons in the literature are primarily interested in grading the relative performance
 58 of different DR algorithms. Though we do not share these goals, many comparative approaches
 59 articulate frameworks and methods with important relationships to our own. We review a selection of
 60 comparison methods, organized in roughly increasing order of the similarity between their goals and
 61 framework to our own.

62 Some evaluation methods for dimensionality reduction lead with intuition, formalizing helpful
 63 heuristics into rigorous ratings. We first reference Lee and Verleysen’s co-ranking matrix⁴, which
 64 measures insertions to (“intrusion events”) and deletions from (“extrusion events”) k -neighborhoods
 65 in the low-dimension vs. high-dimension space. While this measure is non-parametric (with respect to
 66 the data geometry) and thus extremely flexible, it is sensitive only to local structure in the data. Later,
 67 Lee and Verleysen showed that the performance of a DR method closely follows its (a) insensitivity
 68 to norm concentration and (b) plasticity (i.e., the cost function gradient vanishing for distant points)⁷,
 69 leveraging a more geometric perspective to the analysis of DR performance than was typical of the
 70 contemporary literature⁹. Extending this line of thinking, Wang et al¹ recently offered a strictly
 71 empirical investigation of different DR methods in which they consider only the attractive and
 72 repulsive forces of the loss function over varying distance scales. They show that this framework is
 73 sufficient to characterize DR performance and extrapolate robust empirical principles of “good DR
 74 methods” without need of an underlying formalism. While this work offers striking practical insights,
 75 it does not offer an immediate path to describing degrees of divergence between DR methods.

76 We now consider a class of methods we characterize by their tendency to originate in formal,
 77 geometric considerations of manifold learning. Singer and Wu’s vector diffusion distance⁸, an
 78 extension of diffusion embeddings¹⁷, uses local principal component analysis to locally estimate a
 79 connection on the tangent bundle of a data manifold, from which it constructs a lower-dimensional

80 embedding. While the primary goal of their work is to define a manifold learning method, rather than
81 a technique for comparing such methods, their vector diffusion distance explicitly captures a variety
82 of geometric and topological data invariants. These invariants could be used analogously to our use
83 of persistent homology to compare dimensionality reductions of the same dataset. Similarly, Tyagi et
84 al⁵ propose a set of tangent space estimation criteria for manifold sampling (as a function of manifold
85 curvature and intrinsic dimension) that immediately suggest geometric routes of comparison between
86 learned data manifolds. Finally, in more direct alignment with our goals, Sun and Machard propose a
87 geometric theory of manifold learning, which comes equipped with an intrinsic metric^{18,19}. Their
88 approach is firmly grounded in classical information geometry^{20,21}, comparing learned models via
89 the pullback of the Fisher information metric on direct probabilistic encodings of reduced data. While
90 this perspective diverges substantially from our own, our constructions are mutually translatable, and
91 their approach could provide interesting comparison and/or validation.

92 To the best of our knowledge, only two other studies^{2,3} have examined dimensionality reduction
93 through the lens of persistent homology. Both works primarily consider the recovery quality of a
94 known manifold and propose quality metrics derived from persistent homology. Paul and Chalup²
95 compare DR methods while varying manifold complexity, measuring performance by the similarity
96 of pre- and post-DR Betti numbers as a function of sampling density. While their goals differ from
97 ours, much of our comparison also hinges on counts of topological features; however, they did not
98 have access to the topological bootstrap^{14,15} we employ in our study. Rieck and Leitte³ compute
99 the Wasserstein distance between pre- and post-DR persistence diagrams as a metric of embedding
100 quality. While their study is most similar to our own, there are several key differences. First, they
101 operationalize persistent homology with a sublevel-set filtration of the local density function, whereas
102 we use the Vietoris-Rips filtration on pairwise point dissimilarities. Second, they consider reduction
103 of a surface embedded in $D = 3$ reduced to $d = 2$, whereas we consider data initially embedded in
104 $D \sim 10^8$ and reduced to dimensionalities ranging from $d \sim 10^2$ to $d \sim 10^5$. Most importantly, we
105 are able to leverage the topological bootstrap^{14,15} in our work, which was not available at the time of
106 their publication.

107 Finally, we also contrast our analysis of persistence data with a prevalent paradigm in persistent
108 homology applications. Many persistent homology analyses^{6,22–27} (including those above) operationalize
109 the assumption that most topological information lives in a diagram’s most persistent
110 components, treating low-persistence generators as noise. Instead, we follow work in distributed
111 persistence^{28,29} and examine the distributional properties of our persistence diagrams to parse their
112 topological content.

113 Our contributions in the present study are as follows: (1) A flexible framework for the statistical
114 comparison of dimensionality reductions, applicable to any data and dissimilarity measure compatible
115 with a Vietoris-Rips complex; (2) Robust statistical measurement of topological differences between
116 dimension-reduced data; (3) Application to real neuroscience data over a diverse slice of widely used
117 neuroimaging DR algorithms (“brain representation” or “BR”).

118 2 Methods

119 2.1 Brain Data and Brain Representations

120 The data for this study consists of pre-processed resting-state functional MRI data from N=1003
121 Human Connectome Project young adult (HCP-YA)³⁰ subjects. Each subject’s minimally pre-
122 processed data consists of 91,282 spatial “grayordinates” by 1200 time points, giving an embedding
123 dimension of $D \sim 10^8$ in the initial space. We then chose six different brain representations that
124 are both common within the field and showcase the methodological variability of widely adopted
125 techniques. The brain representations we consider can roughly be grouped by their underlying models
126 of brain function. We characterize the first group of methods as seeking to cluster neural activity
127 into spatially contiguous cortical “parcels.” In the parcellation family, we have Yeo’s parcellated
128 networks³¹, Glasser’s multimodal parcellation³², and Schaefer’s local-global parcellation³³. We also
129 sample from a family of low-rank matrix factorization methods that parse non-contiguous networks
130 of functional activity. Independent component analysis (ICA)³⁴, an extension and application
131 refinement of PCA, underlies perhaps the most widely used brain representation in the field³⁵ and
132 thus is represented here. In addition, we consider PROFUMO³⁶, which parses “functional modes” of
133 brain activity from hierarchical Bayesian signal models. Finally, we include the “principal gradient”

134 (or “gradients”)³⁷, a diffusion embedding method that organizes brain function through cortical
 135 geometry.

136 From each brain representation, one or more feature types were computed to reflect the typical use of
 137 brain representations in the neuroimaging literature. The five feature types considered in this work
 138 are as follows: (1) “amplitude,” the average power of the time signal in a given spatial component; (2)
 139 “network matrix (netmat),” the matrix of pairwise Pearson similarities of time courses for each pair of
 140 spatial components; (3) “partial correlation,” the variance-normalized precision matrix; (4) “map,”
 141 the spatial membership weights of a given spatial component in grayordinate space; and (5) “spatial
 142 network matrix”, the matrix of pairwise Pearson similarities of maps for each spatial component.
 143 The decomposition rank, feature types, and number of features for each brain representation is
 144 summarized in Table 1. Note that since subject data are encoded in terms of features, it is the feature
 145 number and **not** the brain representation’s decomposition rank that denotes the dimension d of the
 146 target embedding space in the mapping $\hat{\varphi} : \hat{S} \rightarrow \mathbb{R}^d$. We compare subject-space embeddings using
 147 the pairwise dissimilarities of their points, which we compute as described in the next section.

Representation Name	Decomposition Rank(s) r	Considered Feature Type(s)	Feature Number(s) d
PROFUMO	33	maps, spatial network matrices	$91282 \times 33, \binom{33}{2}$
Dual-regression spatial ICA	15, 25, 50, 100, 200, 300	amplitudes, network matrices, partial network matrices	$r, \binom{r}{2}, \binom{r}{2}$
Glasser parcellation	360	amplitudes, network matrices, partial network matrices	$360, \binom{360}{2}, \binom{360}{2}$
Schaefer parcellation	100, 200, 300, 600	amplitudes, network matrices, partial network matrices	$r, \binom{r}{2}, \binom{r}{2}$
Yeo parcellation	17	amplitudes, network matrices, partial network matrices	$17, \binom{17}{2}, \binom{17}{2}$
Gradient (diffusion embedding)	1, 15, 25, 50, 100, 200, 300	maps	$91282 \times r$

Table 1: The combinations of brain representation, decomposition rank parameters, and feature types investigated in the present work.

148 2.2 Dissimilarity Measures

149 For each brain representation method, decomposition rank within a given representation, and con-
 150 sidered feature type, we compute pairwise distances between all subjects. Each feature type under
 151 consideration is structured either as a vector (maps, amplitudes) or a symmetric positive semidefinite
 152 (SPSD) matrix (network matrices). This bifurcation of data types is echoed in our choice of measures
 153 when computing the dissimilarity between a pair of subjects. In both the vector case and the SPSPD
 154 data case, we ran our analysis using one dissimilarity measure intrinsic to the data type and another
 155 derived from the Pearson correlation. We use Pearson-based dissimilarities in deference to the
 156 ubiquitous use of the Pearson correlation in neuroimaging analyses.

157 We now define the dissimilarity measures we use on vector data. Suppose s_i and s_j are data vectors
 158 in \mathbb{R}^d , and let $\rho(s_i, s_j)$ denote their Pearson correlation. Let $\langle \cdot, \cdot \rangle$ denote the usual inner product on
 159 \mathbb{R}^d . We then define

$$d_{v_1}(s_i, s_j) = 1 - \langle s_i, s_j \rangle^2 \quad (1)$$

$$d_{v_2}(s_i, s_j) = 1 - \rho^2(s_i, s_j), \quad (2)$$

160 assuming the matrix $D_{ij} = \langle s_i, s_j \rangle$ is scaled to have entries in $[0, 1]$. Note that we can interpret d_{v_2}
 161 as approximately the angular distance between the vectors s_i and s_j after each has been centered. We
 162 refer to d_{v_1} as the “inner product divergence” and d_{v_2} as the “Pearson divergence”.

163 In the SPSPD matrix case, we consider the geodesic distance between matrices on the Riemannian
 164 SPD cone³⁸ alongside a (modified) Pearson divergence. The geodesic distance d_{pd_1} on the symmetric
 165 positive definite cone³⁹ is efficiently implemented via the approximate joint diagonalizer⁴⁰, and we
 166 modify the Pearson divergence d_{v_2} for the correlation matrix case by precomposing it with Fisher’s
 167 z-transformation⁴¹ (the inverse hyperbolic tangent function): we write

$$d_{pd_2}(M_i, M_j) = \operatorname{atanh}^* d_{v_2}(m_i, m_j), \quad (3)$$

168 where m_i is the vector of upper-right triangle entries of the symmetric matrix M_i (diagonal excluded).
 169 This precomposition is necessary for correlation matrices, as it normalizes the correlation values
 170 before re-correlating them. In contrast to the vector case, there is no simple comparison to be made
 171 between these two dissimilarity measures.

172 For each combination of brain representation, rank parameter, and feature type shown in Table 1,
 173 we compute pairwise dissimilarity according to both of whichever two measures are relevant. The
 174 subject-pairwise matrix of dissimilarities then forms the Gram matrix used to compute the persistent
 175 homology, as we describe in the next section.

176 2.3 Persistent Homology

177 We compute the Vietoris-Rips persistence¹³ of each Gram matrix (which is obtained as described
 178 above), and we now give a very brief background on Vietoris-Rips persistence. For a thorough
 179 treatment of persistent homology, see Dey and Wang’s text¹³; for a thorough treatment of algebraic
 180 topology preliminaries, see Hatcher’s text⁴².

181 2.3.1 Brief background

182 The topology of a space can be summarized by its *homology groups*, algebraic invariants that describe
 183 its structure. Persistent homology extends the constructions of homology to finite data, delivering a
 184 multiscale and threshold-free estimation of data topology. To compute the persistent homology of a
 185 dataset X , it must first be equipped with a *simplicial structure*: a simplicial complex $K(X)$ is a set
 186 of subsets of X with the property that $\sigma' \in K$ whenever $\sigma' \subset \sigma$ for some $\sigma \in K$, and a *filtration*
 187 is a collection $\{K_t(X)\}$ such that $K_s \subset K_t$ when $s < t$. Homology groups $H_k(K_t(X))$ can be
 188 computed for each simplicial complex, and their *persistence* $\text{PH}_k(X)$ is described by the evolution of
 189 these groups across the filtration. A simple example of a simplicial complex on X is a graph $G(X)$.
 190 If that graph $G(X)$ is weighted, then the family $\{G_r(X)\}$ of graphs obtained from G by thresholding
 191 its edges at weight r is a filtration on X . If $G(X)$ is the graph on X with edge weights given by the
 192 distance between vertices, then the filtration we just described is the *Vietoris-Rips filtration* on X .
 193 Given any dissimilarity matrix d_X , we can assume it is the Gram matrix of some graph $G(X)$ and
 194 compute its Vietoris-Rips persistence $\text{PH}_k(X)$.

195 2.3.2 Topological bootstrap

196 Because it is possible (and, in fact, common) for multiple data elements to define the same homology
 197 generator, bootstrap re-sampling⁴³ is less straightforward in persistent homology than in many other
 198 modes of analysis. However, Reani and Bobrowski recently demonstrated a "topological bootstrap"
 199 method¹⁴ that uses image persistence⁴⁴ to register homology generators found in co-embeddable
 200 spaces. If X, Y can both be embedded into a shared space Z , then the inclusion maps $X \xrightarrow{\iota_X} Z$
 201 and $Y \xrightarrow{\iota_Y} Z$ induce homology maps ι_X^*, ι_Y^* with corresponding filtration maps $\iota_{r,X}^*, \iota_{r,Y}^*$ (assuming
 202 compatible filtrations on each space). A pair of nontrivial elements in $\text{PH}_k(X)$ and $\text{PH}_k(Y)$ is
 203 said to match via Z if $\iota_{r,X}^*$ and $\iota_{r,Y}^*$ map them to the same nontrivial element of $\text{PH}_k(Z)$ for some
 204 filtration value r . For a matched pair, the affinity score α of the match can be computed from ratios
 205 of lengths of intervals in each filtration for which elements in $\text{PH}_k(X)$ and $\text{PH}_k(Y)$ are matched via
 206 Z . We assign $\alpha = 0$ when no match is found and have $\alpha \in (0, 1]$ otherwise.

207 This procedure simplifies substantially in the bootstrapping case; we then have $Z = X$ and $Y =$
 208 $\widehat{X} \subset X$, and we need only check nontrivial elements of $\text{PH}_k(X)$ for matches in $\text{PH}_k(\widehat{X})$. In the
 209 bootstrap setting, Reani and Bobrowski measure the recurrence stability of a nontrivial generator
 210 $\eta \in \text{PH}_k(X)$ by its *prevalence score*

$$\rho(\eta) := \frac{1}{R} \sum_{j=1}^R \alpha(\eta, \widehat{\eta}_j), \quad (4)$$

211 where $\widehat{\eta}_j$ is the match of η in the j^{th} bootstrap. This is just the average affinity (over all bootstraps)
 212 between η and its matches. In the present study, we compute prevalence scores for each generator in
 213 $\text{PH}_1(X)$ for a given subject dissimilarity matrix X .

214 Our implementation⁴⁵ of the topological bootstrap is a mild extension of Garcia-Redondo et al’s
 215 work¹⁵, which efficiently integrates cycle registration with Ripser⁴⁶ and Ripser-image⁴⁷, refines the

216 cycle affinity measures proposed by Reani and Bobrowski, and broadens the conditions under which
 217 topological bootstrapping may be applied.

218 To satisfy the exchangeability criteria necessary for (any) bootstrapping, we also needed to account
 219 for family relationships between subjects in our bootstrap re-samples. Following the approach of
 220 Winkler et al.⁴⁸, we excluded all bootstrap re-samples that placed individuals with the same mother
 221 on different sides of the inclusion/exclusion divide. We conducted cycle registration using $R = 1000$
 222 bootstraps per dataset at 90% re-sampling (without replacement), and we consider $k = 1$ -dimensional
 223 cycle registration in this work.

224 2.3.3 Prevalence-weighted Wasserstein- p distance

225 The space of persistence diagrams is a metric space⁴⁹ under the Wasserstein- p distance, which
 226 previous work³ has used to compare persistence diagrams of different low-dimensional embeddings.
 227 To include statistical information about the stability of homology generators in this comparison, we
 228 define the *prevalence-weighted* Wasserstein- p distance

$$W_p^{(\rho)}(d_1, d_2) := \left(\inf_{\gamma \in \Gamma_{12}} \sum_{x \in d_1} \|x \cdot \rho(x) - \gamma(x) \cdot \rho(\gamma(x))\|_\infty^p \right)^{\frac{1}{p}}. \quad (5)$$

229 Here, d_1 and d_2 are persistence diagrams, Γ_{12} is the set of bijections between d_1 and d_2 , and $\rho(x)$ is
 230 the prevalence of the homology generator x given in 4. This is a simple re-weighting of the usual
 231 Wasserstein distance, modified to incorporate the prevalence score as a summary of per-cycle stability
 232 statistics.

233 2.3.4 The "matched Betti number" $\beta_k^{(\text{matched})}$

234 We also define the "first matched Betti number" $\beta_k^{(\text{matched})}$ as the number of matched cycles (i.e.,
 235 matches with nonzero affinity scores) found in each bootstrapped re-sample. Intuitively, this is a count
 236 of the number of stable generators found in each bootstrap. The Betti numbers of a persistence module
 237 are typically summarized by curves, since each value of a filtration may induce a homology with
 238 a different set of Betti numbers. However, since the topological bootstrap already uses persistence
 239 interval information to find matched cycles and compute their affinity, we may consider $\beta_k^{(\text{matched})}$ as
 240 having "collapsed" these curves via cycle registration. We consider the distribution of bootstrapped
 241 $\beta_k^{(\text{matched})}$ values as a coarse summary of the distributed persistence^{28,29} of a given dissimilarity matrix
 242 d_X .

243 2.4 Study Design

244 In Table 1, we lay out parameter and feature selections considered for each brain representation. For
 245 every representation, bootstrapped persistence is computed for all combinations of feature, parameter,
 246 and dissimilarity measure considered; this gives a total of 90 subject-pairwise dissimilarity matrices
 247 for which we compute $R = 1000$ topological bootstraps. We compute the prevalence-weighted
 248 Wasserstein-2 distance between all pairs of methods and the $\beta_1^{(\text{matched})}$ distributions for each method.
 249 This method-pairwise distance matrix then undergoes Ward hierarchical clustering⁵⁰ to determine
 250 similarity. Our code is publicly available on github.

251 2.4.1 Hypotheses

252 Comparing across feature and metric choices, we expect the SPD matrix geodesic distance to exhibit
 253 less sensitivity to concentration of measure and thus provide greater distinction between brain
 254 representations. We expect that within-feature groupings for map and amplitude will differ very little
 255 between the considered vector dissimilarity measures (equations 1 and 2). For all comparisons, we
 256 expect feature number and type to be more important drivers of differences than decomposition rank.
 257 Finally, within the PROFUMO analysis, we expect that spatial network matrices will be further from
 258 null than spatial maps, where we expect the very high dimensions of the spatial maps to suffer from
 259 concentration of measure.

260 Comparing across different brain representations, we expect to primarily see clustering according to
 261 (approximate) feature number and type, with secondary similarity clusters forming within each given

262 brain representation. We expect our analysis to align with previous results in the literature linking
 263 shared variance in brain representations^{51–54}, the details of which we expand upon in the results
 264 below.

265 3 Results

266 3.1 Persistent homology and dimension reduction

267 We first note several unexpected instances of trivial (or nearly trivial) persistence structure. First,
 268 full correlation matrices generated null H_1 persistence at every decomposition rank in every brain
 269 representation. By contrast, the partial correlation matrices (which is similar by conjugation to
 270 the inverse of the full correlation matrix) have interesting persistence for nearly all feature types,
 271 decomposition ranks, and dissimilarity measures. Additionally, the inner product divergence (1)
 272 generated trivial or almost trivial homology in both maps and amplitudes, across all ranks and
 273 representations; this is not true of the Pearson divergence, which we incorrectly hypothesized would
 274 exhibit similar behavior. A complete list of all methods that exhibited trivial H_1 persistence is given
 275 in Table S1.

276 3.1.1 Effect of embedding dimension

277 Our analysis saw that topological complexity (as measured by H_1 persistence) generally *decreased*
 278 with the number of features considered (Fig S1). Under the geodesic distance, mean prevalence
 279 score increased with feature number; for all other dissimilarity measures, mean prevalence score
 280 was not correlated with feature number (Fig S2). Taken together, these observations suggest that
 281 embeddings in higher dimensions elicit a smaller number of nontrivial H_1 generators which are also
 282 more robust. This runs counter to the consequences we might expect from concentration of measure
 283 in high dimensions, which pushes spaces towards the discrete topology (and thus a higher number of
 284 less stable generators). As expected, we also saw that feature number was a more important driver of
 285 persistence structure than the underlying rank of the decomposition (Fig S3).

286 3.1.2 Persistence vs. prevalence

287 We see evidence further corroborating Reani and Bobrowski’s observation that the most prevalent
 288 cycles are not always the most persistent ones¹⁴. Figure 1 shows a sample persistence diagram in
 289 H_1 (colored by generator prevalence score) and a plot of all persistence-prevalence pairs observed
 290 in this experiment. Both plots demonstrate that cycles with low persistence can still have high
 291 prevalence, suggesting that the topological "noise" may carry meaningful structure in our data. In
 292 addition, we see a substantially richer difference structure between target embeddings when using the
 293 prevalence-weighted Wasserstein-2 distance instead of the classical Wasserstein-2 distance (Figure
 294 S4).

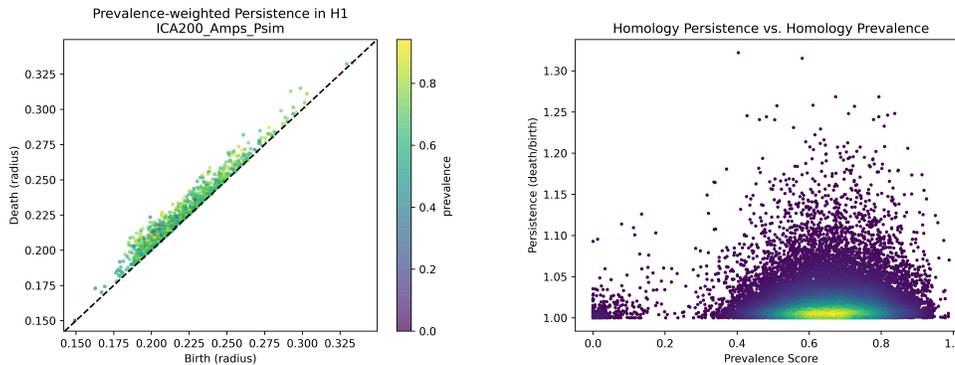


Figure 1: **(Left)** A sample persistence diagram, with color weights given by prevalence score. **(Right)** Persistence versus prevalence across all data collected, colored with a Gaussian kernel density estimator.

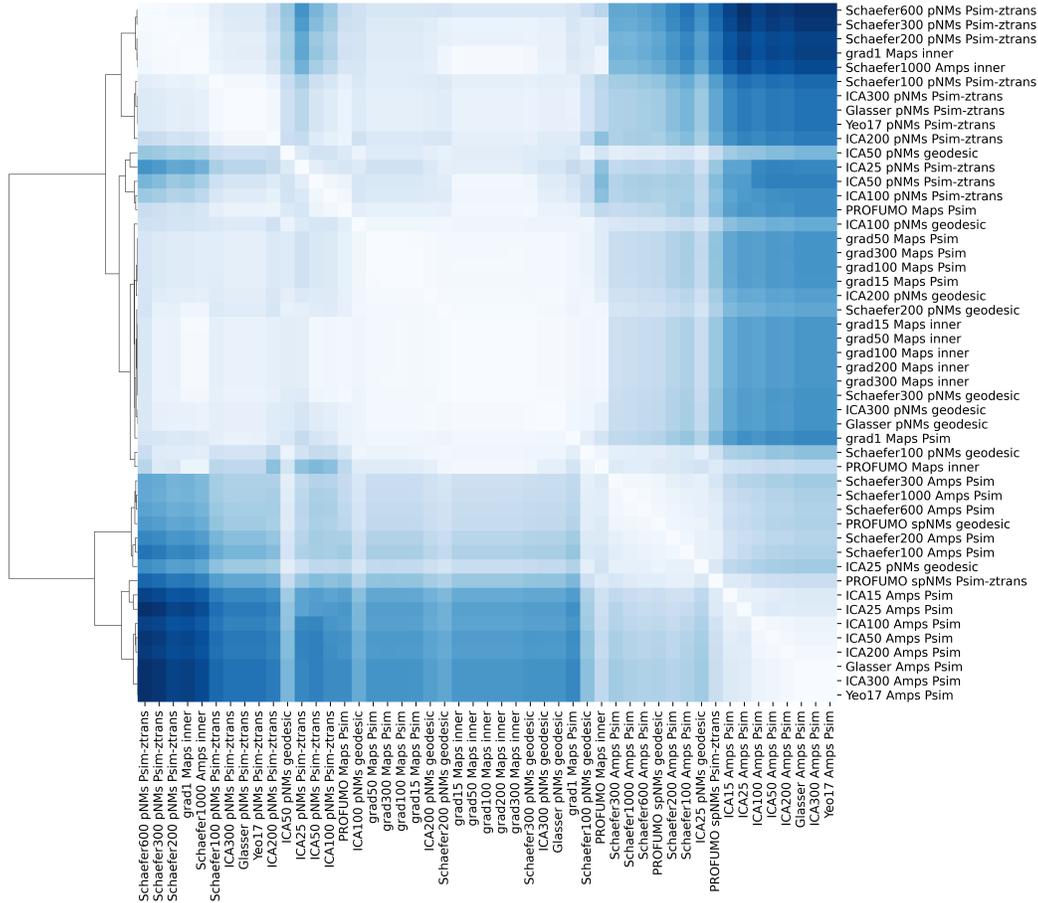


Figure 2: Prevalence-weighted 2-Wasserstein distances between H_1 persistence diagrams for all pairs of methods combinations with nontrivial first homology. Ward hierarchical clustering gives the dendrogram on the left side of the plot, which organizes labels into groups that maximally share variance. Lighter colors denote smaller distances, while darker (blue) colors denote larger ones.

296 The prevalence-weighted Wasserstein distance makes its strongest distinction between amplitudes
 297 and network matrix/spatial map feature types, which form the two main diagonal blocks and highest
 298 dendrogram branches (Fig 2). As hypothesized, this implies that our method distinguishes more
 299 strongly between feature type (and number) than between brain representation type, which forms
 300 the next set of blocks and branches. This is still somewhat surprising, however, because brain
 301 representations differ substantially in terms of whether they are unilateral or bilateral, binary or
 302 weighted, and decomposition rank.

303 We are also surprised to see PROFUMO spatial network matrices in the amplitude block. Both
 304 amplitudes⁵⁵ and spatial network matrices¹² have been shown to be highly sensitive to individual
 305 differences in behavior, but these feature types are interpreted very differently. Amplitudes may be
 306 linked to within-network synchronization⁵⁶, within-network plasticity⁵⁷, or within-network interneu-
 307 ron function⁵⁸, whereas spatial network matrices are indicative of between-network shared brain
 308 regions that may play a role in cross-network integration¹². Both amplitudes and spatial netmats have
 309 higher test-retest reliability (i.e., within-subject stability) than the features in the other block^{36,59}.
 310 Given this context, the clustered blocks of the prevalence-weighted Wasserstein may constitute a
 311 segregation of trait-sensitive (amplitude and spatial network matrix) from state-sensitive (temporal
 312 network matrix) features. This observation highlights the need for an evaluation method that can
 313 detect *which* elements of the persistence module are shared across representations, rather than only
 314 being able to similar topologies of subject similarity.

315 3.3 Computational resources

316 All other computations, including cycle registration, were negligible in cost compared to the com-
317 putation of persistence modules for all bootstraps — roughly 270,000 persistence modules were
318 computed in total. Memory demands remained relatively low ($\leq 50\text{GB}$ per homology computation).
319 Our implementation was embarrassingly parallel on a queue-managed HPC cluster. We estimate that
320 this experiment used approximately 80,000 CPU hours over the course of a month. Computation of
321 image-persistence was the most costly individual step, with each embedded persistence module taking
322 1-3 hours to compute (compared to order of 10 minutes or less for other persistence computations).

323 4 Discussion

324 4.1 Conclusions

325 Our method reveals interesting relationships between dimensionality reductions of resting-state fMRI
326 data. The prevalence-weighted Wasserstein distance distinguishes much more strongly between
327 feature type than dimensionality reduction, potentially segregating trait-sensitive from state-sensitive
328 features. Notably, this distinction holds without regard to choice of dissimilarity measure.

329 Without exception, full network matrices gave rise to trivial PH_1 modules. Persistence modules
330 generated from the inner product divergence (1) were (approximately) trivial as well, in sharp contrast
331 to those generated from the Pearson divergence (2); this suggests that amplitude and spatial map
332 features of brain representations tend to be "mean-dominated," in the sense that per-subject deviations
333 from group-level structures are typically small.

334 In addition, we saw a counterintuitive decrease in persistence "complexity" as a function of increasing
335 embedding dimension, which highlights the difficulties of evaluating dimension reduction in high-
336 dimensional target spaces. We also examined the relationship between persistence and prevalence,
337 finding that the two are largely uncorrelated for our data. Coupled with the stronger distinctions real-
338 ized by the prevalence-weighted Wasserstein-2 distance, we believe that persistence and prevalence
339 may be somewhat complementary as measures of cycle importance.

340 4.2 Limitations

341 Because of the high cost of parameter exploration, dimensionality reduction computation, and
342 topological bootstrapping, only a few dimensionality reduction methods were examined in this work.
343 An extension of this analysis to a wider array of brain representations may be warranted, especially
344 newer methods that derive an explicitly geometric basis for functional activity (e.g., Laplacian
345 eigenvalues⁶⁰).

346 Another important limitation of our work is the very high dimension-to-sample-size ratio ($N \ll d$)
347 of our data. In this regime, it is difficult to ascertain what features we see because of structure in the
348 data and what topological features are products of the curse of dimensionality. This could be partially
349 ameliorated by conducting our analysis over adequately constructed null data and comparing the
350 results, which is beyond the scope of this work.

351 4.3 Future Directions

352 In addition to addressing some of the limitations noted above, we offer several directions for follow-
353 up work on this study. First, we propose a consideration of the per-bootstrap Wasserstein distance
354 between methods; a distributional picture of differences in the endogenous metric of persistence
355 modules could yield important insights. Second, it is possible to repurpose the topological bootstrap
356 to track the addition/deletion of homology components by different brain representation; practically,
357 this is primarily hindered by the lack of a suitable dissimilarity metric between pairs of points under
358 different embeddings. Finding and validating such a metric would be a valuable direction of inquiry.
359 Finally, we wish to suggest an investigation into the theoretical properties of the prevalence-weighted
360 Wasserstein metric.

361 **References**

- 362 1. Wang, Y., Huang, H., Rudin, C. & Shaposhnik, Y. Understanding how dimension reduction
363 tools work: an empirical approach to deciphering t-SNE, UMAP, TriMap, and PaCMAP for data
364 visualization. *J. Mach. Learn. Res.* 22, 201:9129-201:9201 (2022).
- 365 2. Paul, R. & Chalup, S. K. A study on validating non-linear dimensionality reduction using persistent
366 homology. *Pattern Recognition Letters* 100, 160–166 (2017).
- 367 3. Rieck, B. & Leitte, H. Persistent Homology for the Evaluation of Dimensionality Reduction
368 Schemes. *Computer Graphics Forum* 34, 431–440 (2015).
- 369 4. Lee, J. A. & Verleysen, M. Quality assessment of dimensionality reduction: Rank-based criteria.
370 *Neurocomputing* 72, 1431–1443 (2009).
- 371 5. Tyagi, H., Vural, E. & Frossard, P. Tangent space estimation for smooth embeddings of Riemannian
372 manifolds. *Information and Inference: A Journal of the IMA* 2, 69–114 (2013).
- 373 6. Lotz, M. Persistent homology for low-complexity models. *Proceedings of the Royal Society A:
374 Mathematical, Physical and Engineering Sciences* 475, 20190081 (2019).
- 375 7. Lee, J. A. & Verleysen, M. Two key properties of dimensionality reduction methods. in
376 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM) 163–170 (2014).
377 doi:10.1109/CIDM.2014.7008663.
- 378 8. Singer, A. & Wu, H.-T. Vector Diffusion Maps and the Connection Laplacian. *Commun Pure Appl
379 Math* 65, 10.1002/cpa.21395 (2012).
- 380 9. Gracia, A., González, S., Robles, V. & Menasalvas, E. A methodology to compare Dimensionality
381 Reduction algorithms in terms of loss of quality. *Information Sciences* 270, 1–27 (2014).
- 382 10. Amari, S. *Information Geometry and its Applications*. vol. Volume 194 (2016).
- 383 11. Bijsterbosch, J. et al. Challenges and future directions for representations of functional brain
384 organization. *Nature neuroscience* 1–12 (2020) doi:10.1038/s41593-020-00726-z.
- 385 12. Bijsterbosch, J. D., Beckmann, C. F., Woolrich, M. W., Smith, S. M. & Harrison, S. J. The
386 relationship between spatial configuration and functional connectivity of brain regions revisited.
387 *eLife* (2019) doi:10.7554/eLife.44890.001.
- 388 13. Dey, T. K. & Wang, Y. *Computational Topology for Data Analysis*. (Cambridge University Press,
389 2022). doi:10.1017/9781009099950.
- 390 14. Reani, Y. & Bobrowski, O. Cycle Registration in Persistent Homology with Applications in
391 Topological Bootstrap. Preprint at <https://doi.org/10.48550/arXiv.2101.00698> (2021).
- 392 15. García-Redondo, I., Monod, A. & Song, A. Fast Topological Signal Identification and Persistent
393 Cohomological Cycle Matching. Preprint at <https://doi.org/10.48550/arXiv.2209.15446> (2022).
- 394 16. Böhm, J. N., Berens, P. & Kobak, D. Attraction-Repulsion Spectrum in Neighbor Embeddings.
395 arXiv:2007.08902 [cs, stat] (2021).
- 396 17. Coifman, R. R. & Lafon, S. Diffusion maps. *Applied and Computational Harmonic Analysis* 21,
397 5–30 (2006).
- 398 18. Sun, K. & Marchand-Maillet, S. An Information Geometry of Statistical Manifold Learning. in
399 *Proceedings of the 31st International Conference on Machine Learning* 1–9 (PMLR, 2014).
- 400 19. Sun, K. Local measurements of nonlinear embeddings with information geometry. in *Handbook
401 of Statistics* vol. 46 257–281 (Elsevier, 2022).
- 402 20. Amari, S. & Nagaoka, H. *Methods of Information Geometry*. vol. 191 (American Mathematical
403 Society, 2007).
- 404 21. Amari, S.-I., Barndorff-Nielsen, O. E., Kass, R. E., Lauritzen, S. L. & Rao, C. R. *Differential
405 Geometry in Statistical Inference*. *Lecture Notes-Monograph Series* 10, i–240 (1987).

- 406 22. Banman, A.& Ziegelmeier, L. Mind the Gap: A Study in Global Development Through Persistent
407 Homology. in *Research in Computational Topology* (eds. Chambers, E. W., Fasy, B. T.& Ziegelmeier,
408 L.) 125–144 (Springer International Publishing, 2018). doi:10.1007/978-3-319-89593-2_8.
- 409 23. Bhattacharya, S., Ghrist, R.& Kumar, V. Persistent Homology for Path Planning in Uncertain
410 Environments. *IEEE Transactions on Robotics* 31, 578–590 (2015).
- 411 24. Petri, G. et al. Homological scaffolds of brain functional networks. *Journal of the Royal Society*
412 *Interface* (2014) doi:10.1098/rsif.2014.0873.
- 413 25. Lord, L.-D. et al. Insights into Brain Architectures from the Homological Scaffolds of Functional
414 Connectivity Networks. *Frontiers in Systems Neuroscience* 10, 85 (2016).
- 415 26. Fasy, B. T.& Qin, Y. Comparing Distance Metrics on Vectorized Persistence Summaries. 6
416 (2020).
- 417 27. Berry, E., Chen, Y.-C., Cisewski-Kehe, J.& Fasy, B. T. Functional summaries of persistence
418 diagrams. *J Appl. and Comput. Topology* 4, 211–262 (2020).
- 419 28. Solomon, E., Wagner, A.& Bendich, P. From Geometry to Topology: Inverse Theorems for
420 Distributed Persistence. (2021).
- 421 29. Bubenik, P., Hull, M., Patel, D.& Whittle, B. Persistent homology detects curvature. *Inverse*
422 *Problems* 36, 025008 (2020).
- 423 30. Glasser, M. F. et al. The Human Connectome Project’s neuroimaging approach. *Nature*
424 *Neuroscience* 19, 1175–1187 (2016).
- 425 31. Thomas Yeo, B. T. et al. The organization of the human cerebral cortex estimated by intrinsic
426 functional connectivity. *Journal of Neurophysiology* 106, 1125–1165 (2011).
- 427 32. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. *Nature* 536, 171–178
428 (2016).
- 429 33. Schaefer, A. et al. Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic
430 Functional Connectivity MRI. *Cerebral Cortex* 28, 3095–3114 (2018).
- 431 34. Comon, P. Independent Component Analysis, a new concept? *Signal Processing* 36, 287–314
432 (1994).
- 433 35. Nickerson, L. D., Smith, S. M., Öngür, D.& Beckmann, C. F. Using Dual Regression to Investigate
434 Network Shape and Amplitude in Functional Connectivity Analyses. *Frontiers in Neuroscience* 11,
435 115 (2017).
- 436 36. Harrison, S. J. et al. Modelling subject variability in the spatial and temporal characteristics of
437 functional modes. *NeuroImage* 222, 117226 (2020).
- 438 37. Margulies, D. S. et al. Situating the default-mode network along a principal gradient of macroscale
439 cortical organization. *Proceedings of the National Academy of Sciences* 113, 12574–12579 (2016).
- 440 38. Moakher, M. A Differential Geometric Approach to the Geometric Mean of Symmetric Positive-
441 Definite Matrices. *SIAM J. Matrix Anal. & Appl.* 26, 735–747 (2005).
- 442 39. Pennec, X., Fillard, P.& Ayache, N. A Riemannian Framework for Tensor Computing. *Internation-*
443 *ational Journal of Computer Vision* 66, 41–66 (2006).
- 444 40. Congedo, M., Afsari, B., Barachant, A.& Moakher, M. Approximate Joint Diagonalization and
445 Geometric Mean of Symmetric Positive Definite Matrices. *PLOS ONE* 10, e0121423 (2015).
- 446 41. Fisher, R. A. Frequency Distribution of the Values of the Correlation Coefficient in Samples from
447 an Indefinitely Large Population. *Biometrika* 10, 507–521 (1915).
- 448 42. Hatcher, A. *Algebraic Topology*. (Cambridge University Press, 2002).
- 449 43. Abu-Mostafa, Y. S., Magdon-Ismail, M.& Lin, H.-T. *Learning from Data*. (2012).
- 450 44. Cohen-Steiner, D., Edelsbrunner, H., Harer, J.& Morozov, D. Persistent Homology for
451 Kernels, Images, and Cokernels. in *Proceedings of the Twentieth Annual ACM-SIAM Sympos-*
452 *ium on Discrete Algorithms* 1011–1020 (Society for Industrial and Applied Mathematics, 2009).
453 doi:10.1137/1.9781611973068.110.

- 454 45. tyo8/interval-matching-precomp_metric at 67d87792887574195be1376472a9c11f522acc8f.
455 GitHub https://github.com/tyo8/interval-matching-precomp_metric.
- 456 46. Bauer, U. Ripser: efficient computation of Vietoris–Rips persistence barcodes. *J Appl. and*
457 *Comput. Topology* 5, 391–423 (2021).
- 458 47. Bauer, U. & Schmahl, M. Efficient Computation of Image Persistence. Preprint at
459 <https://doi.org/10.48550/arXiv.2201.04170> (2022).
- 460 48. Winkler, A. M., Webster, M. A., Vidaurre, D., Nichols, T. E. & Smith, S. M. Multi-level block
461 permutation. *Neuroimage* 123, 253–268 (2015).
- 462 49. Cohen-Steiner, D., Edelsbrunner, H. & Harer, J. Stability of Persistence Diagrams. *Discrete*
463 *Comput Geom* 37, 103–120 (2007).
- 464 50. Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American*
465 *Statistical Association* 58, 236–244 (1963).
- 466 51. Dadi, K. et al. Benchmarking functional connectome-based predictive models for resting-state
467 fMRI. *NeuroImage* 192, 115–134 (2019).
- 468 52. Botvinik-Nezer, R. et al. Variability in the analysis of a single neuroimaging dataset by many
469 teams. *Nature* 582, 84–88 (2020).
- 470 53. Bijsterbosch, J. D. et al. The relationship between spatial configuration and functional connectivity
471 of brain regions. *eLife* (2018) doi:10.7554/eLife.32992.001.
- 472 54. R, K. et al. Comparison Between Gradients and Parcellations for Functional Connectivity
473 Prediction of Behavior. *Neuroimage* 120044 (2023).
- 474 55. Miller, K. L. et al. Multimodal population brain imaging in the UK Biobank prospective
475 epidemiological study. *Nat Neurosci* 19, 1523–1536 (2016).
- 476 56. Lee, S. et al. Amplitudes of resting-state functional networks - investigation into their correlates
477 and biophysical properties. *Neuroimage* 265, 119779 (2023).
- 478 57. Sydnor, V. J. et al. Intrinsic activity development unfolds along a sensorimotor–association
479 cortical axis in youth. *Nat Neurosci* 26, 638–649 (2023).
- 480 58. Anderson, K. M. et al. Convergent molecular, cellular, and cortical neuroimaging signatures of
481 major depressive disorder. *Proceedings of the National Academy of Sciences* 117, 25138–25149
482 (2020).
- 483 59. Dutt, R. K. et al. Mental health in the UK Biobank: A roadmap to self-report measures and
484 neuroimaging correlates. *Human Brain Mapping* 43, 816–832 (2022).
- 485 60. Pang, J. C. et al. Geometric constraints on human brain function. 2022.10.04.510897 Preprint at
486 <https://doi.org/10.1101/2022.10.04.510897> (2023).