From Persistence to Survival: Hypothesis Testing, Effect Sizes and Vectorisation for Topological Features

Juliette Murris **Bernadette Stolz** **Karsten Borgwardt**

Department of Machine Learning and Systems Biology,
Max Planck Institute of Biochemistry, Martinsried, Germany
{murris, stolz, borgwardt}@biochem.mpg.de

Abstract

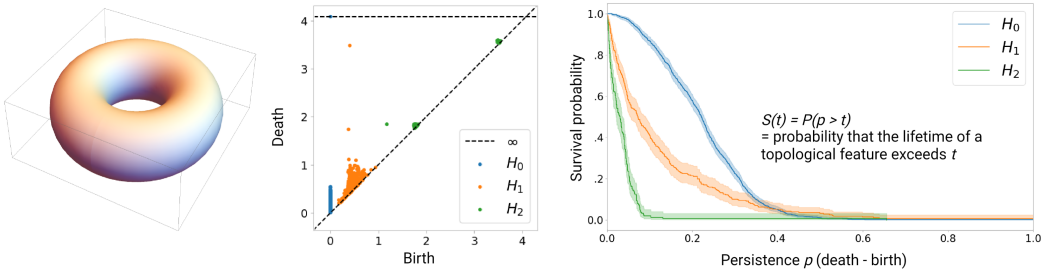
Persistence diagrams are common representations in topological data analysis, but they do not naturally live in a vector space, and the statistical tools developed for comparing them have largely evolved separately from those used for downstream prediction. We introduce `STRAND` (Survival Topological Representation ANalysis of Diagrams), which treats collections of PDs as survival data: each topological feature with persistence $p = d - b$ is a fully observed time-to-event, and groups of diagrams are compared through the persistence survival function $S(t) = \mathbb{P}(p > t)$, the marginal distribution of feature lifetimes. From this single representation we derive (i) a non-parametric two-sample test on collections of diagrams, (ii) interpretable effect sizes, and (iii) a fixed-dimensional feature vector for downstream machine learning, whose integrated deviation is bounded by the 1-Wasserstein distance. We validate calibration and power on synthetic manifolds with controlled topology, show that the vectorisation is competitive across 14 graph and 3D point cloud benchmarks, and apply the method to developmental differences in functional brain connectivity. The implementation of `STRAND` is available at `BorgwardtLab/strand-tda`.

1 Introduction

The shape of data carries information that local statistics miss. A protein's function depends on how amino acids fold into loops and cavities; a brain's organisation manifests in the topology of its functional connections; a drug's efficacy depends on the shape of its molecular graph. Topological data analysis (TDA) [Robins, 2000, Ghrist, 2008, Carlsson, 2009] provides mathematical tools to extract such shape-based descriptors. Persistent homology (PH), the most commonly used method in TDA, tracks topological features (connected components, loops, voids) as they appear and disappear across scales [Edelsbrunner et al., 2002, Zomorodian and Carlsson, 2004]. The output is a persistence diagram (PD), a multiset of birth-death pairs encoding the scales when each feature emerges and vanishes in the data.

Despite their descriptive power, PDs pose two challenges for practitioners [Otter et al., 2017, Chazal and Michel, 2021]. First, given PDs from two conditions, topological difference characterisation remains descriptive rather than confirmatory without principled hypothesis tests [Bubenik and Kim, 2007, Fasy et al., 2014]. Second, PDs are multisets of varying cardinality, incompatible with machine learning algorithms that expect fixed-dimensional inputs [Ali et al., 2023]. Existing work addresses these challenges separately, producing disconnected toolkits. Vectorisation methods [Adams et al., 2017, Bubenik and Dłotko, 2017, Reininghaus et al., 2015, Carrière et al., 2020] enable classification but provide no statistical inference. Conversely, hypothesis testing methods [Robinson and Turner, 2017, Islambekov and Pathirana, 2024, Kwitt et al., 2015, Schrab et al., 2023] yield p-values via permutation or kernel procedures but provide no effect sizes, no localisation of where groups differ,

40th Conference on Neural Information Processing Systems (NeurIPS 2026).



1 Admire the data and decide the maximal homology dimension to investigate

2 Compute the persistence diagram using your favorite TDA library

3 Extract birth, death and persistence values for each feature and estimate the persistence survival probability

Figure 1: From a point cloud (1, here a torus), persistent homology produces a persistence diagram with features in dimensions H_0, H_1 and H_2 (2). Each feature contributes its lifetime $p = d - b$ as a fully observed survival time, and `STRAND` estimates the persistence survival function $S(t) = \mathbb{P}(p > t)$ separately for each homology dimension (3).

and no representation that connects to downstream prediction. A practitioner who wants both classification and statistical comparison must use two different methods operating on two different objects, with the features driving prediction different to the ones tested for significance, breaking the interpretive link between what is predicted and what is reported as significant.

We close this gap with `STRAND` (Survival Topological Representation ANalysis of Diagrams) by mapping PDs to survival data (Figure 1). Each topological feature with persistence value $p_i = d_i - b_i$ is treated as a fully observed time-to-event observation. The obtained persistence survival function $S(t) = \mathbb{P}(p > t)$ enables comparison between groups, yielding a hypothesis test with interpretable effect sizes. Besides, its discretisation yields a fixed-dimensional feature vector for downstream machine learning. Our contributions are:

- **Hypothesis testing for collections of persistence diagrams** (Section 3.2). We introduce $S(t) = \mathbb{P}(p > t)$, the marginal distribution of feature lifetimes, as the central object for comparing collections of diagrams, and we use the log-rank test to compare S_A to S_B between two cohorts.
- **Interpretable, dimension-stratified effect sizes** (Sections 3.2–4.3). Beyond the scalar p-values returned by existing PD inference, `STRAND` reports hazard ratios with bootstrap confidence intervals, median persistence differences, and Kaplan-Meier curves that localise group differences across the persistence axis and across homological dimensions.
- **A vectorisation with an integrated stability bound derived from the same object** (Section 3.3). Discretising $\hat{S}$ on a grid yields a fixed-dimensional vector whose integrated deviation is bounded by the 1-Wasserstein distance (Proposition 3.2). It is competitive with state-of-the-art vectorisations across 14 graph and 3D point cloud benchmarks, and with fewer features than persistence images at matched accuracy.
- **Application to autism connectivity** (Section 4.3). On functional brain connectivity diagrams across three brain atlases, `STRAND` jointly delivers per-dimension hypothesis tests with effect sizes, head-to-head against three established PD-inference baselines.

2 Related Work

Persistent homology. Persistent homology (PH) tracks topological features across scales [Edelsbrunner et al., 2002, Zomorodian and Carlsson, 2004, Carlsson, 2009]. Given a filtration $\mathcal{F} = \{K_\epsilon\}_{\epsilon \geq 0}$ of nested simplicial complexes (e.g., Vietoris-Rips on point clouds, degree-based on graphs), PH records when connected components (H_0), loops (H_1), and voids (H_2) appear and disappear. The output is a persistence diagram (PD) $\mathcal{D} = \{(b_i, d_i)\}_{i=1}^n$ with persistence $p_i = d_i - b_i$. PDs are stable under the Wasserstein and bottleneck distances [Cohen-Steiner et al., 2005, Skraba and Turner, 2020].

Vectorisation. Vectorisation methods map PDs to fixed-dimensional vectors compatible with machine learning. Persistence Images [Adams et al., 2017] and Persistence Landscapes [Bubenik and

Dłotko, 2017] produce stable representations via Gaussian smoothing and piecewise-linear functions respectively; Betti curves [Giusti et al., 2015] and the broader Persistence Curves family [Chung and Lawson, 2022] aggregate feature counts under a weight-function parametrisation, and filtration curves [O'Bray et al., 2021] generalise the Betti curve on graphs by tracking any graph descriptor along an edge filtration (see Appendix I.4 for a geometric comparison). Kernel methods [Reininghaus et al., 2015, Carriere et al., 2017, Le and Yamada, 2018] embed diagrams into reproducing kernel Hilbert space for kernel machines. Neural approaches learn representations directly [Carrière et al., 2020, Xin et al., 2023, Verma et al., 2024]; see Ali et al. [2023] for a survey. None of these methods provides hypothesis testing with interpretable effect sizes.

Statistical inference. Robinson and Turner [2017] introduced a permutation test on Wasserstein distances, and Islambekov and Pathirana [2024] extended this to vectorised Betti functions. Thomas et al. [2023] test a single diagram against a noise model with a statistic that integrates the log-count of features exceeding each lifetime. Confidence-set methods [Fasy et al., 2014] target single-diagram inference. Kernel-based two-sample tests apply maximum mean discrepancy [Gretton et al., 2012] via persistence-specific kernels [Kwitt et al., 2015, Reininghaus et al., 2015, Carriere et al., 2017, Schrab et al., 2023]. Persistence kernels can in principle do both, but in practice the two are split, with a kernel for classification, and a separate permutation or MMD test for inference. Across this landscape, every method yields a p-value but none yields an interpretable effect size or localises group differences to specific filtration scales (Appendix B.4, Table 6).

Survival analysis. Survival analysis models time-to-event data [Clark et al., 2003]. The survival function $S(t) = \mathbb{P}(T > t)$ gives the probability that a random time T exceeds t, and the Kaplan-Meier estimator [Kaplan and Meier, 1958] provides a non-parametric estimate from observed event times. The log-rank test [Mantel, 1966] compares two survival distributions under $H^{\text{null}} : S_A = S_B$, with the median survival times and hazard ratios quantifying effect sizes.

Further background on TDA and survival analysis appears in Appendix B.

3 Survival Topological Representation ANalysis of Diagrams (STRAND)

We introduce STRAND, that maps persistence diagrams (PDs) to survival data via the *persistence survival function*. The same object underlies complementary capabilities: a non-parametric two-sample test for comparing topological structure between groups (Section 3.2) along with effect sizes, and a fixed-dimensional vectorisation for downstream machine learning (Section 3.3). We first define the central object and state how it depends on the diagram (Section 3.1).

3.1 From Persistence to Survival

Let $\mathcal{D} = \{(b_i, d_i)\}_{i=1}^n$ denote a PD, where each topological feature i appears at filtration value b_i and disappears at d_i. The *persistence value* $p_i = d_i - b_i$ measures the lifetime of feature i across the filtration. STRAND treats each p_i as a fully observed time-to-event and the multiset $\{p_i\}_{i=1}^n$ as a survival sample, which in this paper shall be static and fully observed. The correspondence with classical survival analysis is detailed in Appendix A. We define the **persistence survival function**

$$S(t) = \mathbb{P}(p > t), \tag{1}$$

the probability that a randomly chosen feature persists beyond filtration scale t. Equivalently, S is determined by the cumulative hazard $\Lambda(t) = -\log S(t)$, which records the cumulative rate of feature disappearance up to t. S takes values in $[0, 1]$ and is the natural object for vectorisation (Section 3.3), while Λ is additive across non-overlapping intervals, which is what makes the log-rank test of Section 3.2 a sum of contributions over event times.

Non-parametric estimation. Given a PD with distinct ordered persistence values $t_1 < t_2 < \cdots < t_m$, let n_j be the number of features at risk just before t_j (i.e., features with $p_i \geq t_j$) and e_j the number of events at t_j. The Kaplan–Meier estimator of the survival function and the Nelson–Aalen estimator of the cumulative hazard are

$$\hat{S}(t) = \prod_{t_j \leq t} \frac{n_j - e_j}{n_j}, \qquad \hat{\Lambda}(t) = \sum_{t_j \leq t} \frac{e_j}{n_j}, \tag{2}$$

with variance estimated by Greenwood's formula $\widehat{\mathrm{Var}}(\hat{S}(t)) = \hat{S}(t)^2 \sum_{t_j \le t} e_j/(n_j(n_j - e_j))$ [Kaplan and Meier, 1958, Aalen, 1978]. PDs additionally exhibit a delayed-entry structure, with features only observable between birth and death. This is accommodated through left truncation, with the at-risk set redefined as $n_j = \#\{i : b_i < t_j \le d_i\}$. While Eq. (2) corresponds to a marginal model on persistence values, the left-truncated form corresponds to a model on death times conditional on entry at birth.

Estimands and sampling model. We distinguish two sampling levels. *Within a single diagram*, the multiset $\{p_i\}_{i=1}^n$ defines an empirical persistence measure $\hat{\mu}_{\mathcal{D}} = \frac{1}{n}\sum_{i=1}^n \delta_{p_i}$, and $\hat{S}$ is its empirical survival function. *Across diagrams sampled from a distribution P on $\mathcal{D}$*, the pooled estimand is

$$S_P(t) = \mathbb{E}_{D\sim P}\big[\,\mathbb{P}_{p|D}(p > t)\,\big], \tag{3}$$

the marginal under the mixture of sampling a diagram from P then a feature uniformly from that diagram, well-defined whenever $\mathbb{E}_P[|D|] < \infty$. The test of Section 3.2 compares S_{P_A} to S_{P_B}. Pooling the features of a collection weights each diagram by its size, so the pooled estimator targets $\tilde{S}_P(t) = \mathbb{E}_P[\,|D|\,S_D(t)\,]/\mathbb{E}_P[\,|D|\,]$ with $S_D(t) = \mathbb{P}_{p|D}(p > t)$, which equals Eq. (3) when diagram size carries no information about lifetimes. Weighting each feature by $1/|D|$ recovers Eq. (3) [Williamson et al., 2003]. The survival function also determines every integrable functional of persistence, including the mean and total persistence (Appendix I.1).

Stability. We state how the survival vectorisation $\phi_S(\,\cdot\,;\mathbf{t}) : \mathcal{D} \to [0,1]^k$, which evaluates $\hat{S}$ on a grid $\mathbf{t} = (t_1, \ldots, t_k)$, depends on the diagram. The statements hold for a grid fixed in advance, as under global normalisation, and under local normalisation the grid itself depends on the diagram. Proofs are in Appendix C.

Theorem 3.1 (Pointwise bound)**.** *Let $D = \{(b_i, d_i)\}_{i=1}^n$ and $D' = \{(b'_j, d'_j)\}_{j=1}^m$ be PDs with $n, m \ge 1$ off-diagonal points. Then*

$$\|\phi_S(D;\mathbf{t}) - \phi_S(D';\mathbf{t})\|_\infty \;\le\; \frac{4\,W_1(D,D')}{\delta_\Delta \cdot \max(n,m)} \;+\; \frac{2\,W_1(D,D')}{\delta \cdot \min(n,m)}, \tag{4}$$

where $W_1(D,D')$ is the 1-Wasserstein distance, δ is the minimum nonzero gap between matched persistence values, and δ_Δ is the minimum persistence among features matched to the diagonal under an optimal W_1 transport, a term being zero when its set is empty. If $n = m$ and some optimal transport γ^* matches every point of D off the diagonal with $p_i = p'_{\gamma^*(i)}$, then $\phi_S(D;\mathbf{t}) = \phi_S(D';\mathbf{t})$.

Each matched pair costs at most $W_1(D,D')$, so δ, $\delta_\Delta \le 2\,W_1(D,D')$ whenever they are defined. The second term is therefore at least $1/\min(n,m)$ as soon as two matched persistence values differ, and Eq. (4) does not vanish as $W_1(D,D') \to 0$. This reflects two discontinuities of evaluation on a grid, *i)* a persistence value crossing a grid point moves one coordinate by $1/n$, and *ii)* a feature of arbitrarily small persistence changes the normalisation from $1/n$ to $1/(n+1)$. The integrated survival function has neither.

Proposition 3.2 (Integrated stability)**.** *Let $N_D(t) = \#\{i : p_i > t\}$. For any PDs D and D', $\int_0^\infty |N_D(t) - N_{D'}(t)|\,dt \le 2\,W_1(D,D')$. In particular, if $|D| = |D'| = n$,*

$$\int_0^\infty \big|\hat{S}_D(t) - \hat{S}_{D'}(t)\big|\,dt \;\le\; \frac{2\,W_1(D,D')}{n}. \tag{5}$$

Proposition 3.3 (Eventual equality on a fixed grid)**.** *Let D have n off-diagonal points, no grid point equal to a persistence value of D, and $\eta = \min\big(\min_i p_i,\ \min_{i,l} |p_i - t_l|\big) > 0$. Then $\phi_S(D';\mathbf{t}) = \phi_S(D;\mathbf{t})$ for every PD D' with $|D'| = n$ and $W_1(D,D') < \eta/2$.*

Relation to existing summaries. The count $N_D(t) = n\,\hat{S}_D(t)$ of Proposition 3.2 is the stable rank of the barcode under the standard contour [Chachólski and Riihimäki, 2020] and the ALPS statistic of Thomas et al. [2023], defined for H_0, integrates $\log N_D(t)$ over t. $\hat{S}$ is distinct from the Betti curve $\beta(t) = \#\{i : b_i \le t \le d_i\}$, and neither determines the other. As an example, the diagrams $D = \{(0,3),(1,2)\}$ and $D' = \{(0,2),(1,3)\}$ have identical Betti curves at $W_1(D,D') = 2$, yet their lifetimes are $\{3,1\}$ and $\{2,2\}$. Conversely, $\hat{S}$ is invariant to translations $(b,d) \mapsto (b+c, d+c)$ while β is not. Geometrically, $\beta(t)$ counts points in the quadrant $\{b \le t \le d\}$, whose corner slides along the diagonal, while $\hat{S}(t)$ counts points above the line $d - b = t$ (Figure 7, Appendix I.4).

3.2 Hypothesis Testing on Persistence-Lifetime Distributions

The persistence survival function enables hypothesis testing on the lifetime distributions of collections of persistence diagrams, with calibrated Type I error and effect sizes that go beyond a scalar p-value.

Two-sample test. Given two groups of PDs $\mathcal{D}_A = \{D_1^A, \ldots, D_{n_A}^A\}$ and $\mathcal{D}_B = \{D_1^B, \ldots, D_{n_B}^B\}$, `STRAND` tests $H^{\text{null}} : S_{P_A}(t) = S_{P_B}(t)$ for all t against $H^{\text{alt.}} : S_{P_A}(t) \neq S_{P_B}(t)$ for some t. We proceed as follows:

1. Pool persistence values within each group: $T_A = \{p_i : (b_i, d_i) \in D_j^A,\ j = 1, \ldots, n_A\}$, and analogously for T_B;
2. Apply the log-rank test to compare $S_{P_A}(t)$ versus $S_{P_B}(t)$.

At each distinct event time t_j, let n_{Aj}, n_{Bj} denote the at-risk counts and d_{Aj}, d_{Bj} the event counts in groups A and B, with $n_j = n_{Aj} + n_{Bj}$ and $d_j = d_{Aj} + d_{Bj}$. Under H^{null}, the conditional expectation and hypergeometric variance of d_{Aj} given the margins are

$$E_{Aj} = \frac{n_{Aj}\, d_j}{n_j}, \qquad V_j = \frac{n_{Aj}\, n_{Bj}\, d_j\, (n_j - d_j)}{n_j^2\, (n_j - 1)}. \tag{6}$$

The log-rank numerator $U = \sum_j (d_{Aj} - E_{A_j})$ compares observed with expected events in group A. Its classical variance $\sum_j V_j$ treats pooled persistence values as independent, which they are not, as features of one diagram are coupled together in H_0 and by homological constraints in H_k, $k \geq 1$. We therefore treat the diagram as the sampling unit and estimate the variance of U from diagram-level score residuals [Jung and Jeong, 2003]. For diagram i with group indicator $z_i \in \{0, 1\}$ (1 for A) and features $k = 1, \ldots, n_i$, let $Y_{ik}(t) = \mathbf{1}\{p_{ik} \geq t\}$, $N_{ik}(t) = \mathbf{1}\{p_{ik} \leq t\}$, $\bar{z}(t)$ the fraction of at-risk features in group A, and $\hat{\Lambda}$ the pooled Nelson–Aalen estimator. Then

$$\hat{U}_i = \sum_{k=1}^{n_i} \int_0^\infty \big(z_i - \bar{z}(t)\big)\big\{dN_{ik}(t) - Y_{ik}(t)\, d\hat{\Lambda}(t)\big\}, \qquad \chi^2_{\text{cl}} = \frac{U^2}{\sum_i \hat{U}_i^2} \xrightarrow{H^{\text{null}}} \chi^2_1 \tag{7}$$

as the number of independently sampled diagrams grows, whatever the dependence between features of the same diagram. With few diagrams per group we use a permutation test that reassigns group labels to whole diagrams, which is exact under exchangeability of diagrams; with equal group sizes and a two-sided statistic it cannot reach $p < 0.05$ below four diagrams per group. `STRAND` extends to $k > 2$ groups through the k-sample log-rank statistic with $k - 1$ degrees of freedom.

Effect sizes. Under proportional hazards, the hazard ratio HR such that $\lambda_B(t) = \text{HR} \cdot \lambda_A(t)$ summarises the relative rate of feature disappearance between groups; $\widehat{\text{HR}}$ is estimated non-parametrically from the at-risk and event counts in (6), with confidence intervals from a 1,000-sample bootstrap over diagrams. We assess proportional hazards via Schoenfeld residuals [Schoenfeld, 1982]. When the proportional hazards assumption does not hold, we report instead the median persistence difference $\Delta_{\text{med}} = m_A - m_B, m_g = \inf\{t : S_g(t) \leq 1/2\}$, which remains a valid non-parametric effect-size summary without distributional assumptions. Appendix I.5 discusses procedures for the three regimes (PH holds, monotone deviation, non-monotone deviation).

Statistical power. Topological complexity governs statistical power: manifolds with richer homology produce more features, increasing power at fixed effect size.

Proposition 3.4 (Power under proportional hazards)**.** *Assume (i) the persistence values pooled within each group form an i.i.d. sample from S_{P_A}, S_{P_B} respectively, (ii) the proportional hazards model $\lambda_B = \text{HR} \cdot \lambda_A$ holds, and (iii) equal allocation $n_A = n_B$. Then the asymptotic power of the log-rank test at level α against the alternative* HR *is*

$$1 - \Phi\left(z_{1-\alpha/2} - \frac{|\log \text{HR}|\, \sqrt{n_{\text{eff}}}}{2}\right) + o(1),$$

where $n_{\text{eff}} := |PD^A| + |PD^B|$ is the total number of persistence features across both groups [Schoenfeld, 1981].

Assumption *(i)* ignores dependence between features of the same diagram. Under the clustered variance of Eq. (7), n_{eff} is replaced approximately by $n_{\text{eff}}/\{1+(\bar{n}-1)\bar{\rho}\}$, with $\bar{n}$ the mean diagram size and $\bar{\rho}$ the average within-diagram correlation, so that when $\bar{\rho} > 0$ power is limited by the number of diagrams (and not the number of features).

For point clouds of N points sampled from a d-dimensional manifold $\mathcal{M}$ at density ρ, $\mathbb{E}[|PD_k|] = O(\rho^{d-k}\,\beta_k(\mathcal{M}))$, where β_k is the k-th Betti number [Giusti et al., 2015]. For graphs with n vertices and edge density p, degree-based filtrations yield $\mathbb{E}[|PD_0|] = n-1$ and $\mathbb{E}[|PD_1|] = O(pn^2 - n)$, providing concrete guidance for power calculations. The same scaling that makes features pool large also raises the question of whether these produce statistically significant but practically negligible differences; we address this in Appendix I.6.

3.3 STRAND vectorisation

Because S is a one-dimensional summary, it admits a natural fixed-dimensional discretization. STRAND's vectorisation uses the same $\hat{S}$ that drives the hypothesis test, so the features fed to a classifier are exactly the quantities tested for significance.

Grid discretization. Given a PD $\mathcal{D}$ with estimated survival function $\hat{S}_{\mathcal{D}}$, we evaluate at g equally-spaced points spanning the diagram's own observed persistence range (local normalisation, Appendix G.2):

$$t_k = t_{\min} + \frac{k-1}{g-1}(t_{\max} - t_{\min}), \quad k = 1, \ldots, g. \tag{8}$$

The STRAND vector for homology dimension h is $\mathbf{v}^{(h)} = \left[\hat{S}^{(h)}_{\mathcal{D}}(t_1), \ldots, \hat{S}^{(h)}_{\mathcal{D}}(t_g)\right] \in \mathbb{R}^g$. For PDs with multiple homology dimensions we concatenate: $\mathbf{v} = \left[\mathbf{v}^{(0)} \,\|\, \mathbf{v}^{(1)} \,\|\, \cdots\right] \in \mathbb{R}^{g\cdot|\mathcal{H}|}$. The grid resolution g controls the trade-off between representational fidelity and feature dimensionality.

Variants and their estimands. The variants estimate distinct functionals of the diagram-generating process:

- STRAND **(standard, primary)** uses the Kaplan-Meier estimator of Eq. (2) on persistence values, targeting the marginal $S(t) = \mathbb{P}(p > t)$. Each feature is treated as entering at time zero. This variant produces $g \cdot |\mathcal{H}|$ features and is the default in all experiments unless otherwise stated.
- STRAND**-F (birth features, augmented primary)** augments the standard vector with a birth survival curve, $\hat{S}_{\text{birth}}(t) = \mathbb{P}(b > t)$, adding birth information to the lifetime marginal. The final vector is $\mathbf{v}_{\texttt{STRAND-}F} = [\mathbf{v}_{\text{persistence}} \,\|\, \mathbf{v}_{\text{birth}}] \in \mathbb{R}^{2g\cdot|\mathcal{H}|}$. STRAND-F is well-defined only when birth times vary; for H_0 from sublevel-set or Vietoris-Rips filtrations on connected complexes, births are identically zero, and STRAND-F reduces to STRAND.
- STRAND**-B (birth entry, alternative)** uses the left-truncated Kaplan-Meier estimator, with features entering the risk set at their birth time. The estimand is the conditional death-time survival $S_B(t) = \mathbb{P}(d > t \mid b < t)$, related to a hazard $\lambda_d(t) = \lim_{\Delta\to 0}\mathbb{P}(d \in [t, t+\Delta) \mid d \geq t,\, b < t)/\Delta$ on the death-time axis. The marginal formulation is appropriate when persistence itself is the structural object of interest, which is typical in TDA classification (see Appendix I.2 for empirical analysis of why the conditional estimand underperforms in classification).

Complexity. For a PD with n points, the Kaplan-Meier estimator requires $O(n \log n)$ time (dominated by sorting persistence values); grid evaluation is $O(g)$ via binary search. For $|\mathcal{H}|$ homology dimensions, total per-diagram complexity is $O(|\mathcal{H}| \cdot (n \log n + g))$. STRAND's practical advantage stems from low output dimensionality: with $g = 25$ and two homology dimensions, STRAND produces 58 features (25 grid values and four scalar summaries per dimension: the median persistence, its confidence bounds and the feature count), or 112 for STRAND-F, which adds the birth grid and two birth scalars.

4 Experiments

We evaluate STRAND on three fronts: (i) statistical validation on synthetic manifolds with controlled topology ; (ii) classification benchmarks on graph and 3D point cloud datasets, showing the survival

Table 1: Type I error at nominal $\alpha = 0.05$ (200 points per cloud, 1,000 replications per cell, Wilson 95% CIs). Classical: the variance $\sum_j V_j$ of Eq. (6); Clustered: Eq. (7); Permutation: labels reassigned to whole diagrams ($B = 1{,}000$).

Setting	Dim	Classical $\hat{\alpha}$ [95% CI]	Clustered $\hat{\alpha}$ [95% CI]	Permutation $\hat{\alpha}$ [95% CI]
Parametric distributions	–	0.056 [0.044, 0.069]	–	–
Torus (4D)	H_0	0.671 [0.641, 0.699]	0.054 [0.042, 0.070]	0.052 [0.040, 0.068]
	H_1	0.219 [0.194, 0.246]	0.066 [0.052, 0.083]	0.051 [0.039, 0.066]
Klein bottle (4D)	H_0	0.266 [0.240, 0.294]	0.059 [0.046, 0.075]	0.049 [0.037, 0.064]
	H_1	0.057 [0.044, 0.073]	0.055 [0.042, 0.071]	0.047 [0.036, 0.062]
Sphere, cube noise	H_0	0.505 [0.474, 0.536]	0.051 [0.039, 0.066]	0.052 [0.040, 0.068]
	H_1	0.058 [0.045, 0.074]	0.054 [0.042, 0.070]	0.044 [0.033, 0.059]
Sphere, plane noise	H_0	0.446 [0.415, 0.477]	0.037 [0.027, 0.051]	0.059 [0.046, 0.075]
	H_1	0.104 [0.087, 0.124]	0.058 [0.045, 0.074]	0.058 [0.045, 0.074]
Sphere, line noise	H_0	0.414 [0.384, 0.445]	0.022 [0.015, 0.033]	0.052 [0.040, 0.068]
	H_1	0.088 [0.072, 0.107]	0.055 [0.042, 0.071]	0.043 [0.032, 0.057]
Sphere, Laplace noise	H_0	0.548 [0.517, 0.579]	0.029 [0.020, 0.041]	0.049 [0.037, 0.064]
	H_1	0.094 [0.077, 0.114]	0.065 [0.051, 0.082]	0.046 [0.035, 0.061]

vectorisation is competitive with state-of-the-art representations; and (iii) an application to autism brain connectivity, where `STRAND` quantifies developmental topological differences.

4.1 Hypothesis Testing Validation

We validate `STRAND`'s inference capabilities through controlled experiments on synthetic manifolds, assessing proper Type I error control and power as a function of the number of diagrams per group.

Setup. Following Stolz [2023], we generate point clouds containing mixtures of signal points (sampled from known manifolds) and noise points. We evaluate on three manifolds: a 4D torus, a 4D Klein bottle, and a 3D sphere with various noise configurations. For each setting, we vary the signal probability $p_{\text{signal}} \in [0, 1]$ across 50 values with 10 independent realizations each, yielding 500 datasets per manifold. Full simulation details are provided in Appendix D.

Type I Error Control. We test under H^{null} by drawing two groups of diagrams from the same signal mixture. Table 1 reports rejection rates with the classical variance $\sum_j V_j$, with the clustered variance of Eq. (7), and with a permutation test that reassigns labels to whole diagrams. The classical test is invalid in this design , because it counts features and not diagrams. The clustered test stays within $[0.022, 0.066]$ across every manifold and both dimensions, and the permutation test, which is exact under exchangeability of diagrams, is nominal throughout. The agreement validates the clustered test, while it does not establish that features are independent. Single-diagram contrasts, where the feature is both the sampling unit and the unit of interest and the classical variance is the appropriate one, are calibrated as well (Appendix D).

Sample Efficiency. Table 2 reports power on simulated 4D torus data when two groups of n diagrams are compared. Both the clustered and the permutation test reach full power from five diagrams per group . Below that, no procedure gives both validity and power, and the clustered variance is asymptotic in the number of diagrams, while the permutation test is limited by its discreteness, since with equal groups its smallest attainable two-sided p-value is $2/\binom{2n}{n}$, which exceeds 0.05 for $n \leq 3$. We therefore state sample efficiency from $n = 5$ and recommend the exact permutation test below that.

Power to Detect Topological Differences. We assess `STRAND`'s ability to distinguish signal (manifold) from noise features using log-rank tests on pooled persistence values. Figure 2 illustrates results on the 4D torus across the signal probability spectrum. Across all homology dimensions, `STRAND` achieves strong discrimination, with p-values consistently below 0.05 throughout the signal probability range. The median survival difference shows interpretable patterns: at low signal probability, genuine topological features persist substantially longer than noise; this gap narrows as signal

Table 2: Power of the clustered STRAND test on simulated 4D torus data at $\alpha = 0.05$, for two groups of n diagrams (1,000 replications; permutation test with $B = 1{,}000$; cells give clustered / permutation). Point-cloud rows compare two single diagrams of N points, where only the classical test is defined.

Experiment	Size	H_0	H_1
Point cloud	$N = 30$	0.80	0.32
	$N = 50$	1.00	0.67
	$N = 100$	1.00	0.78
	$N = 150$	1.00	0.81
Collection	$n = 5$	1.00 / 1.00	1.00 / 1.00
	$n = 10$	1.00 / 1.00	1.00 / 1.00
	$n = 20$	1.00 / 1.00	1.00 / 1.00

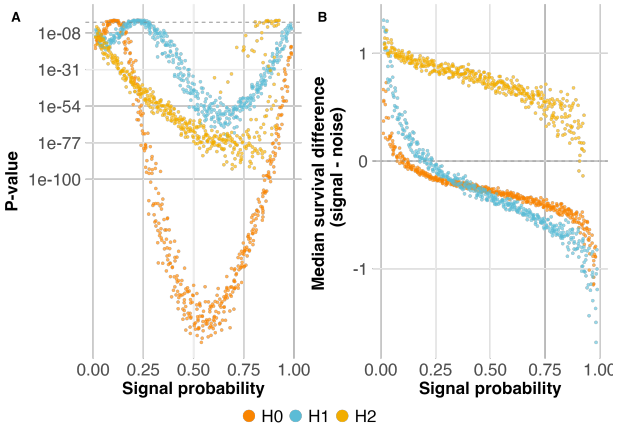


Figure 2: Hypothesis testing validation on 4D torus (500 simulations). A: Log-rank p-values comparing signal vs. noise survival distributions, dashed line indicates $\alpha = 0.05$. B: Median survival difference quantifies effect size.

dominates the point cloud. Power as a function of hazard ratio on the 4D torus follows the expected V-shape, with a rejection rate of 0.070 at $\mathrm{HR} = 1$ for the clustered test (see Appendix F).

4.2 Classification Benchmark

While hypothesis testing is STRAND's primary contribution, the same survival function provides a fixed-dimensional vector representation suitable for downstream machine learning. We benchmark this vectorisation across graph and 3D point cloud classification at scale.

Setup. We evaluate on ten graph datasets from the TU Dortmund collection [Morris et al., 2020], and on four 3D point cloud datasets from the flooder package [Graf et al., 2026] (datasets description in Appendix G, Table 12). Following Carrière et al. [2020], we compute PDs using degree-based filtration [Horn et al., 2022] in dimensions H_0 and H_1 for each graph. For point clouds, we use the flood complex [Graf et al., 2026] with H_0, H_1, and H_2. We use Support Vector Machine and Random Forest classifiers across all vectorisations, and SVM only for the two kernel baselines, which supply a Gram matrix rather than a feature vector, with stratified 10-fold cross-validation (graphs) or the 10 pre-defined splits provided by Graf et al. [2026] (point clouds).

We compare STRAND against a comprehensive set of vectorisation and kernel baselines: Persistence Images (PI) [Adams et al., 2017], Persistence Landscapes (PL) [Bubenik and Dłotko, 2017], Betti Curves (BC) [Giusti et al., 2015], Silhouette and Entropy summaries from the GUDHI library, PersLay (fixed, used as a vectorizer for fair comparison) [Carrière et al., 2020], and two persistence kernels: Sliced Wasserstein Kernel (SWK) [Carriere et al., 2017] and Persistence Scale Space Kernel (PSSK) [Reininghaus et al., 2015]. Hyperparameters were selected by grid search on the training portion of each split (Appendix G.2).

Results. Table 3 shows that on TU Dortmund graphs, all methods cluster within a narrow accuracy range (58–63% on average), reflecting the limited topological signal of degree filtration on these small graphs [Morris et al., 2020]. STRAND, SWK, and PI lead with 61.5%, 61.6%, 61.3%; pairwise differences are not statistically significant (Wilcoxon signed-rank, all $p > 0.5$). On the geometrically richer point cloud benchmarks, STRAND achieves the highest average accuracy (79.5% vs. 76.7% for PI, +2.8pp), with the largest margins on the most complex multi-class datasets (MCB +3.0pp, ModelNet10 +5.6pp). To our knowledge, this is the first systematic application of TDA vectorisations to point clouds at this scale. Across the 14 datasets, STRAND achieves the best accuracy on 7 and is within 1–5pp of the best baseline on the remaining 7 (see Appendix I.2 for STRAND variants discussion).

Table 3: Classification accuracy (%) with SVM for every method. STRAND reports the best variant per dataset. **Bold**: best per row; underline: second best.

Dataset	STRAND	PI	PL	BC	Silh.	Entr.	PLay	SWK	PSSK
TU Dortmund graphs (degree filtration, H_0/H_1; 10-fold CV)									
COX2	78.5	78.7	78.2	78.8	78.2	78.2	**78.9**	78.2	78.2
DD	71.3	75.0	72.1	**75.4**	69.4	69.4	74.8	**75.4**	74.3
DHFR	67.6	64.3	61.0	62.6	60.9	63.4	61.2	**68.8**	61.0
ENZYMES	**27.0**	24.3	25.2	22.0	25.7	21.0	23.4	21.1	21.2
IMDB-B	64.2	62.5	63.7	63.4	**64.4**	64.2	63.5	63.7	63.6
IMDB-M	**41.3**	40.3	39.2	38.4	39.8	37.1	40.0	40.1	39.2
MUTAG	77.9	**78.8**	73.8	78.4	68.3	65.8	75.2	**78.8**	66.5
NCI1	62.6	63.0	60.4	62.0	57.1	57.1	63.1	**63.2**	61.2
PROTEINS	66.4	71.0	70.6	70.0	66.4	67.3	70.8	**71.3**	60.6
PTC-MR	**58.5**	55.3	54.7	53.6	56.8	56.3	56.2	55.7	57.6
Average	61.5	61.3	59.9	60.5	58.7	58.0	60.7	**61.6**	58.3
3D point cloud datasets (flood complex, $H_0/H_1/H_2$; 10 pre-defined splits)									
SwissCheese	**100.0**	97.9	88.1	98.5	98.7	98.8	–	–	–
Corals	**70.8**	70.0	66.0	70.2	61.9	64.4	63.1	–	–
MCB	**81.9**	78.9	80.3	69.0	46.6	51.4	–	–	–
ModelNet10	**65.5**	59.9	55.3	52.5	45.3	55.7	–	–	–
Average	**79.5**	76.7	72.4	72.5	63.1	67.6	63.1	–	–

PI = Persistence Images, PL = Persistence Landscapes, BC = Betti Curves, PLay = PersLay (fixed), SWK = Sliced Wasserstein Kernel, PSSK = Persistence Scale Space Kernel. Kernel methods (SWK, PSSK) are computationally infeasible on the larger flooder datasets due to $O(N^2)$ Gram matrices and per-pair complexity scaling with diagram size.

Computational Efficiency. STRAND's advantage beyond accuracy is feature parsimony and computational tractability. Table 4 compares feature dimensionality and SVM runtime on the TU Dortmund average. STRAND uses 58–112 features versus 5,000 for PI on a 50×50 grid, a 45–86$\times$ reduction with no loss in accuracy. The advantage compounds on flooder datasets where kernel methods become infeasible: SWK and PSSK require $O(N^2)$ Gram matrices with per-pair cost scaling in diagram size. On ModelNet10 ($N = 4{,}899$, flood complexes on 250,000-point clouds), SWK is estimated at $\sim 10^4$ hours versus 258 seconds for STRAND. PersLay's TensorFlow fitting similarly scales prohibitively. STRAND's cost is $O(n \log n)$ per diagram with no pairwise comparisons.

Table 4: Computational efficiency on the TU Dortmund average. Features: output dimensionality. Time: SVM pipeline (single AMD EPYC 9654 CPU node).

Method	Features	Time (s)
STRAND ($g = 25$)	58	6.1
STRAND-F ($g = 25$)	112	18.9
PI	5,000	188.8
PL	500	15.4
BC	100	6.9

4.3 Case study on functional connectivity data in autism

Prior topological work on developmental connectomes has established that connectome topology carries developmentally meaningful signal [Gracia-Tabuenca et al., 2023]. We illustrate STRAND on characterising topological differences between children's and adults' brain networks.

Setup. We use the Autism Brain Imaging Data Exchange (ABIDE I) [Heinsfeld et al., 2018], restricted to typically developing controls, and contrast children (age < 13, $n = 130$) with adults (age ≥ 18, $n = 110$). For each subject and atlas, we build the Pearson correlation matrix between ROI resting-state time series[1] and convert it to a distance matrix with $d_{ij} = \sqrt{2(1 - \rho_{ij})}$, where ρ_{ij} is the signed Pearson correlation. Persistence diagrams in H_0 and H_1 are computed with Ripser from the Vietoris–Rips filtration of each distance matrix, truncated at the 0.90 quantile of that subject's pairwise distances; features still alive at the threshold are discarded. Subjects are retained if mean framewise displacement is at most 0.2 mm, at most 50% of frames exceed 0.5 mm, at least 100

[1] Pearson correlation is one of several functional-connectivity measures. The choice can influence network structure [Bullmore and Bassett, 2011, Smith et al., 2011, Zhou et al., 2009]. We use Pearson correlation throughout for compatibility with the developmental-connectivity literature we anchor against [Gracia-Tabuenca et al., 2023].

Table 5: ABIDE developmental contrast (children < 13 vs. adults ≥ 18) across three atlases. p_{cl}: clustered log-rank with each subject's diagram as a cluster; p_{perm}: permutation over subjects ($B = 1{,}000$, floor 0.001); HR: inverse-size-weighted hazard ratio with a 95% CI from 1,000 subject-bootstrap resamples (unweighted values and design effects in Table 22). Comparators use $B = 1{,}000$ label permutations. **Bold**: $p < 0.05$; †: $p_{\mathrm{cl}} < 0.05/6$.

Dim	Atlas	p_{cl}	p_{perm}	HR [95% CI]	R&T	MMD-SWK	MMD-PSSK
H_0	Schaefer-400	$\mathbf{< 0.001}^{\dagger}$	$\mathbf{< 0.001}$	0.815 [0.731, 0.909]	**0.007**	0.182	**0.011**
	CC200	0.743	0.743	0.982 [0.881, 1.099]	0.492	0.523	0.482
	AAL90	0.504	0.498	0.967 [0.878, 1.072]	0.268	**0.025**	0.602
H_1	Schaefer-400	$\mathbf{< 0.001}^{\dagger}$	$\mathbf{< 0.001}$	1.148 [1.110, 1.191]	**0.008**	**0.001**	**0.001**
	CC200	$\mathbf{< 0.001}^{\dagger}$	$\mathbf{< 0.001}$	1.103 [1.060, 1.147]	**0.008**	**0.008**	**0.001**
	AAL90	$\mathbf{< 0.001}^{\dagger}$	**0.002**	1.092 [1.041, 1.148]	**0.019**	**0.010**	**0.001**

time points are available, and ROI loss stays within an atlas-specific cap. To assess sensitivity to ROI partition, we replicate under three brain parcellations of differing granularity, AAL90 (90 ROIs) [Tzourio-Mazoyer et al., 2002], CC200 (200 ROIs) [Craddock et al., 2012], and Schaefer-400 (400 ROIs) [Schaefer et al., 2018]. We then test whether the children's and adults' collections of PDs differ, and characterise the difference if one is detected.

Inference and comparison methods. For each parcellation and homological dimension, `STRAND` returns the clustered log-rank p-value of Eq. (7), with each subject's diagram as a cluster and sites pooled, a permutation p-value that reassigns labels to whole diagrams ($B = 1{,}000$), and a hazard ratio (HR) with a 95% confidence interval from 1,000 bootstrap resamples of subjects. We compare against three established PD-inference baselines on the same diagrams, each with $B = 1{,}000$ label permutations: the Wasserstein-permutation test of Robinson and Turner [2017] (R&T), and kernel maximum mean discrepancy (MMD) with the Sliced Wasserstein Kernel [Carriere et al., 2017] (MMD-SWK) and the Persistence Scale Space Kernel [Reininghaus et al., 2015] (MMD-PSSK).

Findings. Table 5 reports results across atlases and methods. `STRAND` detects a developmental difference at H_1 on every parcellation, with HR > 1 in every case, indicating that loops in children's diagrams have shorter persistence than in adults'. The HR magnitudes give a quantitative read of the effect, with the increasing magnitude at finer parcellations consistent with finer ROI partitions resolving more of the distinguishing topological structure (further results are reported in Appendix H). At H_0, the result is parcellation-dependent. `STRAND` detects the difference at Schaefer-400 (HR $= 0.815$, CI $[0.731, 0.909]$, $p_{\mathrm{cl}} = 3.1 \times 10^{-4}$) but not at coarser parcellations, and two of the three comparator methods (R&T, MMD-PSSK) replicate this pattern, indicating the H_0 signal lives at the parcellation grid resolution rather than at a methodological choice.

5 Conclusion

`STRAND` is designed for settings where a binary p-value is not enough, where practitioners need to know how much two cohorts differ. Such settings are widespread, e.g., drug discovery, where compound libraries differ topologically in ring systems (H_1) and binding pockets (H_2) by quantifiable amounts, biomedical imaging, where confirmatory studies require sample-size calculations grounded in effect-size estimates, and the analysis of neural network activation manifolds [Gardinazzi et al., 2024, Bhan et al., 2025], where topological characterisations of training and fine-tuning regimes are an emerging direction. In each of these, `STRAND` replaces the current dichotomy between vectorisation pipelines (predictive but silent on inference) and pairwise tests (inferential but silent on magnitude and direction) with a single object that does both. The main limitations of `STRAND` are that inference needs the clustered variance and at least five diagrams per group, that the lifetime marginal is blind to diagonal translations, and that the stability guarantee is integrated rather than pointwise (Appendix I.7). Natural extensions of our work include attribute-aware filtrations on graphs [Uddin et al., 2026] and exploration of cases when features still alive at the cutoff are deemed right-censored in a truncated filtration.

Acknowledgments and Disclosure of Funding

The authors are very grateful to the reviewers for their fruitful feedback. This work received no external funding. Computations were performed on the HPCL cluster at the Max Planck Computing and Data Facility. The authors declare no competing interests.

References

Odd Aalen. Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pages 701–726, 1978.

Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. In *Journal of Machine Learning Research*, volume 18, pages 1–35, 2017.

Dashti Ali, Aras Asaad, Maria-Jose Jimenez, Vidit Nanda, Eduardo Paluzo-Hidalgo, and Manuel Soriano-Trigueros. A survey of vectorization methods in topological data analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12):14069–14080, 2023.

Milan Bhan, Jean-Noel Vittaut, Nicolas Chesneau, Sarath Chandar, and Marie-Jeanne Lesot. Did i faithfully say what i thought? bridging the gap between neural activity and self-explanations in large language models. *arXiv preprint arXiv:2506.09277*, 2025.

Cristian Bodnar, Fabrizio Frasca, Nina Otter, Yuguang Wang, Pietro Lio, Guido F Montufar, and Michael Bronstein. Weisfeiler and lehman go cellular: Cw networks. *Advances in neural information processing systems*, 34:2625–2640, 2021a.

Cristian Bodnar, Fabrizio Frasca, Yuguang Wang, Nina Otter, Guido F Montufar, Pietro Lió, and Michael Bronstein. Weisfeiler and lehman go topological: Message passing simplicial networks. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 1026–1037. PMLR, 18–24 Jul 2021b. URL https://proceedings.mlr.press/v139/bodnar21a.html.

Peter Bubenik and Paweł Dłotko. A persistence landscapes toolbox for topological statistics. *Journal of Symbolic Computation*, 78:91–114, 2017.

Peter Bubenik and Peter T Kim. A statistical approach to persistent homology. 2007.

Edward T Bullmore and Danielle S Bassett. Brain graphs: graphical models of the human brain connectome. *Annual review of clinical psychology*, 7(1):113–140, 2011.

Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

Mathieu Carrière and Ulrich Bauer. On the Metric Distortion of Embedding Persistence Diagrams into Separable Hilbert Spaces. In Gill Barequet and Yusu Wang, editors, *35th International Symposium on Computational Geometry (SoCG 2019)*, volume 129 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 21:1–21:15, Dagstuhl, Germany, 2019. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-104-7. doi: 10.4230/LIPIcs.SoCG.2019.21. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.SoCG.2019.21.

Mathieu Carriere, Marco Cuturi, and Steve Oudot. Sliced wasserstein kernel for persistence diagrams. In *International conference on machine learning*, pages 664–673. PMLR, 2017.

Mathieu Carrière, Frédéric Chazal, Yuichi Ike, Théo Lacombe, Martin Royer, and Yuhei Umeda. Perslay: A neural network layer for persistence diagrams and new graph topological signatures. In *International Conference on Artificial Intelligence and Statistics*, pages 2786–2796. PMLR, 2020.

Wojciech Chachólski and Henri Riihimäki. Metrics and Stabilization in One Parameter Persistence. *SIAM Journal on Applied Algebra and Geometry*, 4(1):69–98, 2020. doi: 10.1137/19M1243932. URL https://doi.org/10.1137/19M1243932.

Frédéric Chazal and Bertrand Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence*, 4:667963, 2021.

Yu-Min Chung and Austin Lawson. Persistence curves: A canonical framework for summarizing persistence diagrams. *Advances in Computational Mathematics*, 48(1):6, 2022.

Taane G Clark, Michael J Bradburn, Sharon B Love, and Douglas G Altman. Survival analysis part i: basic concepts and first analyses. *British journal of cancer*, 89(2):232–238, 2003.

David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. In *Proceedings of the twenty-first annual symposium on Computational geometry*, pages 263–271, 2005.

David R Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

R Cameron Craddock, G Andrew James, Paul E Holtzheimer III, Xiaoping P Hu, and Helen S Mayberg. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33(8):1914–1928, 2012.

Edelsbrunner, Letscher, and Zomorodian. Topological persistence and simplification. *Discrete & computational geometry*, 28(4):511–533, 2002.

Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Confidence sets for persistence diagrams. 2014.

Yuri Gardinazzi, Karthik Viswanathan, Giada Panerai, Alessio Ansuini, Alberto Cazzaniga, and Matteo Biagetti. Persistent topological features in large language models. *arXiv preprint arXiv:2410.11042*, 2024.

Robert Ghrist. Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.

Chad Giusti, Eva Pastalkova, Carina Curto, and Vladimir Itskov. Clique topology reveals intrinsic geometric structure in neural correlations. *Proceedings of the National Academy of Sciences*, 112 (44):13455–13460, 2015.

Zeus Gracia-Tabuenca, Juan Carlos Díaz-Patiño, Isaac Arelio-Ríos, Martha Beatriz Moreno-García, Fernando A Barrios, and Sarael Alcauter. Development of the functional connectome topology in adolescence: evidence from topological data analysis. *ENeuro*, 10(2), 2023.

Florian Graf, Paolo Pellizzoni, Martin Uray, Stefan Huber, and Roland Kwitt. The flood complex: Large-scale persistent homology on millions of points. *Advances in Neural Information Processing Systems*, 38:118356–118389, 2026.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The journal of machine learning research*, 13(1):723–773, 2012.

Mustafa Hajij, Lennart Bastian, Sarah Osentoski, Hardik Kabaria, John L Davenport, Sheik Dawood, Balaji Cherukuri, Joseph G Kocheemoolayil, Nastaran Shahmansouri, Adrian Lew, et al. Copresheaf topological neural networks: A generalized deep learning framework. *arXiv preprint arXiv:2505.21251*, 2025.

Elizabeth A Handorf, Marc C Smaldone, Sujana Movva, and Nandita Mitra. Analysis of survival data with nonproportional hazards: A comparison of propensity-score-weighted methods. *Biometrical journal*, 66(1):202200099, 2024.

Anibal Sólon Heinsfeld, Alexandre Rosa Franco, R Cameron Craddock, Augusto Buchweitz, and Felipe Meneguzzi. Identification of autism spectrum disorder using deep learning and the abide dataset. *NeuroImage: clinical*, 17:16–23, 2018.

Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep learning with topological signatures. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Max Horn, Edward De Brouwer, Michael Moor, Yves Moreau, Bastian Rieck, and Karsten Borgwardt. Topological graph neural networks. In *International Conference on Learning Representations (ICLR)*, 2022. URL https://openreview.net/pdf?id=oxxUMeFwEHd.

Johanna Immonen, Amauri Souza, and Vikas Garg. Going beyond persistent homology using persistent homology. *Advances in neural information processing systems*, 36:63150–63173, 2023.

Hemant Ishwaran, Udaya B Kogalur, Eugene H Blackstone, and Michael S Lauer. Random survival forests. 2008.

Umar Islambekov and Hasani Pathirana. Vector summaries of persistence diagrams for permutation-based hypothesis testing. *Foundations of Data Science*, 6(1):41–61, 2024.

Sin-Ho Jung and Jong-Hyeon Jeong. Rank tests for clustered survival data. *Lifetime data analysis*, 9 (1):21–33, 2003.

Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

Jared L Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. Deepsurv: personalized treatment recommender system using a cox proportional hazards deep neural network. *BMC medical research methodology*, 18(1):24, 2018.

Roland Kwitt, Stefan Huber, Marc Niethammer, Weili Lin, and Ulrich Bauer. Statistical topological data analysis-a kernel perspective. *Advances in neural information processing systems*, 28, 2015.

Tam Le and Makoto Yamada. Persistence fisher kernel: A riemannian manifold kernel for persistence diagrams. *Advances in neural information processing systems*, 31, 2018.

Changhee Lee, William Zame, Jinsung Yoon, and Mihaela Van Der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Nathan Mantel. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*, 50(3):163–170, 1966.

Michael Moor, Max Horn, Bastian Rieck, and Karsten Borgwardt. Topological autoencoders. In *International conference on machine learning*, pages 7045–7054. PMLR, 2020.

Christopher Morris, Nils M Kriege, Franka Bause, Kristian Kersting, Petra Mutzel, and Marion Neumann. Tudataset: A collection of benchmark datasets for learning with graphs. *arXiv preprint arXiv:2007.08663*, 2020.

Juliette Murris, Olivier Bouaziz, Michal Jakubczak, Sandrine Katsahian, and Audrey Lavenu. Random survival forests for the analysis of recurrent events for right-censored data, with or without a terminal event. *BMC Medical Research Methodology*, 25(1):262, 2025.

Leslie O'Bray, Bastian Rieck, and Karsten Borgwardt. Filtration curves for graph representation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 1267–1275, 2021.

Nina Otter, Mason A Porter, Ulrike Tillmann, Peter Grindrod, and Heather A Harrington. A roadmap for the computation of persistent homology. *EPJ Data Science*, 6(1):17, 2017.

Theodore Papamarkou, Tolga Birdal, Michael Bronstein, Gunnar Carlsson, Justin Curry, Yue Gao, Mustafa Hajij, Roland Kwitt, Pietro Lio, Paolo Di Lorenzo, et al. Position: Topological deep learning is the new frontier for relational learning. *Proceedings of machine learning research*, 235: 39529, 2024.

Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4741–4748, 2015.

Vanessa Robins. *Computational topology at multiple resolutions: foundations and applications to fractals and dynamics*. PhD thesis, University of Colorado at Boulder, 2000.

Andrew Robinson and Katharine Turner. Hypothesis testing for topological data analysis. *Journal of Applied and Computational Topology*, 1(2):241–261, 2017.

Alexander Schaefer, Ru Kong, Evan M Gordon, Timothy O Laumann, Xi-Nian Zuo, Avram J Holmes, Simon B Eickhoff, and BT Thomas Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral cortex*, 28(9):3095–3114, 2018.

David Schoenfeld. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*, 68(1):316–319, 1981.

David Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1): 239–241, 1982.

Antonin Schrab, Ilmun Kim, Mélisande Albert, Béatrice Laurent, Benjamin Guedj, and Arthur Gretton. Mmd aggregated two-sample test. *Journal of Machine Learning Research*, 24(194):1–81, 2023.

Primoz Skraba and Katharine Turner. Wasserstein stability for persistence diagrams. *arXiv preprint arXiv:2006.16824*, 2020.

Stephen M Smith, Karla L Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F Beckmann, Thomas E Nichols, Joseph D Ramsey, and Mark W Woolrich. Network modelling methods for fmri. *Neuroimage*, 54(2):875–891, 2011.

Bernadette J Stolz. Outlier-robust subsampling techniques for persistent homology. *Journal of Machine Learning Research*, 24(90):1–35, 2023.

Andrew M Thomas, Peter A Crozier, Yuchen Xu, and David S Matteson. Feature detection and hypothesis testing for extremely noisy nanoparticle images using topological data analysis. *Technometrics*, 65(4):590–603, 2023.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Octave Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

Md Joshem Uddin, Astrit Tola, Cuneyt Akcora, and Baris Coskunuzer. Topoformer: Topology meets attention for graph learning. In *International Conference on Learning Representations*, volume 2026, pages 69300–69325, 2026.

Yogesh Verma, Amauri H Souza, and Vikas Garg. Topological neural networks go persistent, equivariant, and continuous. In *ICML*, 2024.

Zhiyu Wang, Arian Jamasb, Mustafa Hajij, Alex Morehead, Luke Braithwaite, and Pietro Liò. Topotein: Topological deep learning for protein representation learning. *arXiv preprint arXiv:2509.03885*, 2025.

Ronald L Wasserstein and Nicole A Lazar. The asa statement on p-values: context, process, and purpose, 2016.

John M Williamson, Somnath Datta, and Glen A Satten. Marginal analyses of clustered data when cluster size is informative. *Biometrics*, 59(1):36–42, 2003.

Cheng Xin, Soham Mukherjee, Shreyas N Samaga, and Tamal K Dey. Gril: A 2-parameter persistence based vectorization for machine learning. In *Topological, Algebraic and Geometric Learning Workshops 2023*, pages 313–333. PMLR, 2023.

Cheng Xin, Fan Xu, Xin Ding, Jie Gao, and Jiaxin Ding. Toping: Topologically interpretable graph learning via persistent rationale filtration. In *ICML*, 2025.

Dongli Zhou, Wesley K Thompson, and Greg Siegle. Matlab toolbox for functional connectivity. *Neuroimage*, 47(4):1590–1607, 2009.

Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. In *Proceedings of the twentieth annual symposium on Computational geometry*, pages 347–356, 2004.

A Bridging TDA and survival analysis

Our approach reads the birth-death lifecycle of topological features constitutes a time-to-event process amenable to survival analysis. To illustrate this mapping, consider a simple example:

Example 3.1 (Circle with noise). Suppose we have a point cloud sampled from a circle with added Gaussian noise, as per Figure 3. As we increase the filtration parameter ϵ from 0, we observe the following feature lifecycle:

- $\epsilon = 0.1$: Many small connected components appear (birth events);
- $\epsilon = 0.3$: Components begin merging; the main circular structure emerges as a 1-dimensional loop (loop birth);
- $\epsilon = 0.5$: Most noise components die (death events for components); the main loop persists;
- $\epsilon = 0.8$: The circular loop finally dies when the "hole" gets filled (loop death).

In survival analysis terminology, each connected component and loop is a "subject" entering the study at its birth time and experiencing a "death event" when it disappears. The analogy is one of structure, not of data: a feature's lifetime is a fixed quantity computed from the filtration, observed in full for every feature, whereas a patient's survival time unfolds in calendar time and may be censored. Figure 3 illustrates this conceptual mapping. Topological features become subjects with birth times as entry points and death times as events, while filtration parameters provide temporal structure. This mapping enables statistical comparison between subgroups (signal and noise in TDA, or treatment groups in clinical studies) using established survival methodologies.

This enables both discovery of signal-noise structure and classification of individual features using well-established statistical methodology.

B Further background and related work

B.1 Persistent homology

Simplicial complexes. A *simplicial complex* K is a collection of simplices (vertices, edges, triangles, tetrahedra, and their higher-dimensional analogues) closed under taking faces: if $\sigma \in K$ and $\tau \subseteq \sigma$, then $\tau \in K$. The k-skeleton of K consists of all simplices of dimension at most k. Formally, a k-simplex is the convex hull of $k+1$ affinely independent points; for example, a 0-simplex is a vertex, a 1-simplex is an edge, and a 2-simplex is a triangle.

A *filtration* is a nested sequence of simplicial complexes $\emptyset = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K$, where simplices are added incrementally. The filtration value $f(\sigma)$ of a simplex σ is the smallest index i such that $\sigma \in K_i$. Common filtrations include:

- *Vietoris-Rips filtration*: Given a point cloud $X \subset \mathbb{R}^d$ and scale parameter ϵ, the Rips complex $\mathrm{VR}_\epsilon(X)$ includes a simplex $\{x_0, \ldots, x_k\}$ if all pairwise distances satisfy $\|x_i - x_j\| \leq \epsilon$.
- *Čech filtration*: The Čech complex $\check{C}_\epsilon(X)$ includes a simplex if the corresponding ϵ-balls have non-empty intersection. This is homotopy equivalent to the union of balls but more expensive to compute.
- *Alpha filtration*: A subset of the Delaunay triangulation, providing a sparser representation with the same persistent homology as the Čech complex.

Persistence diagrams from point cloud data. Given a point cloud $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^d$, persistent homology tracks how topological features evolve across scales. The standard pipeline proceeds as follows:

1. *Filtration construction*: Build a filtered simplicial complex, typically the Vietoris-Rips filtration $\{\mathrm{VR}_\epsilon(X)\}_{\epsilon \geq 0}$.
2. *Boundary matrix computation*: Represent the boundary operator $\partial_k : C_k \to C_{k-1}$ as a matrix, where C_k is the vector space of k-chains.
3. *Matrix reduction*: Apply the standard reduction algorithm [Edelsbrunner et al., 2002] to compute the persistence pairs, identifying which simplices create (birth) and destroy (death) homology classes.

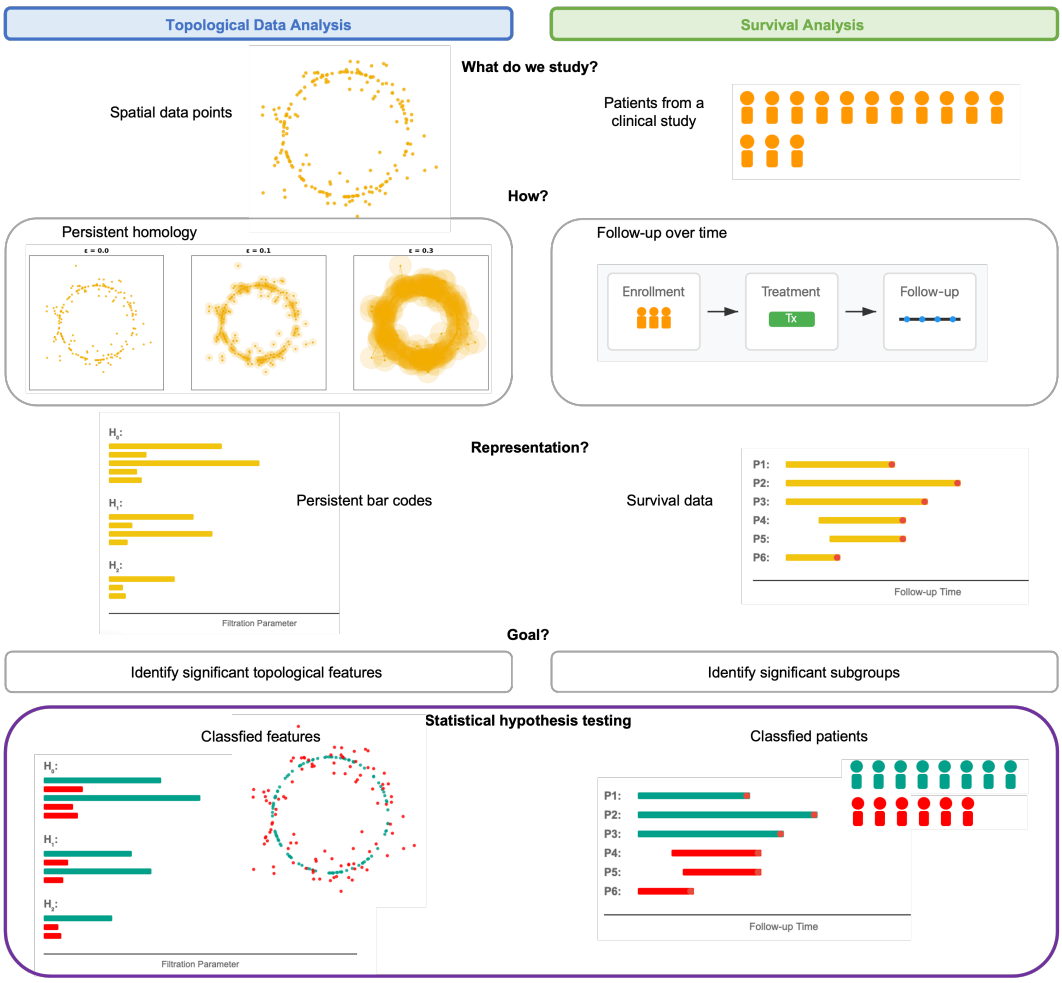


Figure 3: Bridging topological data and survival analysis fields with `STRAND`. Our approach operates on spatial data containing genuine topological structure and random noise artifacts. Through persistent homology, we track how topological features emerge and disappear across increasing filtration parameters. Simultaneously, in survival analysis, patients progress through clinical enrollment, treatment, and follow-up phases, generating survival timelines. In both fields, the goal is usually to identify significant populations, either features in TDA, or responders in clinical studies.

4. *Diagram extraction*: Output the PD $\mathcal{D} = \{(b_i, d_i)\}$, where each point represents a feature born at filtration value b_i and dying at d_i.

The resulting diagram is stable: if two point clouds X and X' satisfy $d_H(X, X') \leq \epsilon$ (Hausdorff distance), then $W_\infty(\mathcal{D}_X, \mathcal{D}_{X'}) \leq \epsilon$ [Cohen-Steiner et al., 2005].

Persistence diagrams from graphs. For graph data $G = (V, E)$, several filtration strategies produce meaningful PDs:

- *Degree filtration* [Hofer et al., 2017]: Vertices enter the filtration in order of increasing degree, and edges appear when both endpoints are present. This captures hierarchical connectivity: hub vertices (high degree) appear late, while peripheral vertices appear early.
- *Heat kernel signature (HKS) filtration* [Carrière et al., 2020]: Vertices are ordered by the diagonal of the heat kernel $H_t = e^{-tL}$, where L is the graph Laplacian. This requires eigendecomposition and depends on the diffusion time t.
- *Sublevel set filtration on node attributes*: When vertices have scalar attributes $f : V \to \mathbb{R}$, the sublevel set filtration adds vertices in order of $f(v)$ and edges when both endpoints are present.

For graph filtrations, the H_0 features correspond to connected components merging as the filtration progresses, while H_1 features correspond to cycles forming and being filled in. The degree filtration used in this work is parameter-free and has been shown to capture chemically meaningful patterns in molecular graphs [Horn et al., 2022].

B.2 Topological Deep Learning

Recent work integrates topological information into neural architectures. Higher-order message passing on simplicial and cell complexes [Bodnar et al., 2021a,b, Hajij et al., 2025] generalizes GNNs beyond pairwise interactions. Persistence-augmented GNNs [Horn et al., 2022, Immonen et al., 2023] incorporate topological features as auxiliary inputs, with applications to molecular property prediction [Xin et al., 2025, Wang et al., 2025]. Deep learning with topological signatures [Hofer et al., 2017] and topological autoencoders [Moor et al., 2020] demonstrate the value of persistence-based representations; see Papamarkou et al. [2024] for a position paper on the field. `STRAND` complements these architectures: its features can augment any GNN as graph-level descriptors while additionally providing statistical inference capabilities.

B.3 Survival Analysis in Machine Learning

Survival analysis models time-to-event data using the Kaplan-Meier estimator [Kaplan and Meier, 1958] for nonparametric survival estimation and the Cox proportional hazards model [Cox, 1972] for covariate analysis. The Cox model assumes the hazard function takes the form $h(t|x) = h_0(t)\exp(\beta^\top x)$, where $h_0(t)$ is a baseline hazard and β are regression coefficients. The proportional hazards assumption states that the hazard ratio between any two individuals is constant over time.

Modern extensions include random survival forests [Ishwaran et al., 2008, Murris et al., 2025], which aggregate survival trees using random feature selection and bootstrap aggregation, and neural approaches such as DeepSurv [Katzman et al., 2018], which replaces the linear predictor with a neural network, and DeepHit [Lee et al., 2018], which directly models the probability mass function of event times.

The log-rank test [Mantel, 1966] is the standard nonparametric test for comparing survival distributions between groups. Under the null hypothesis of equal survival functions, the test statistic

$$\chi^2 = \frac{\left(\sum_j (O_{1j} - E_{1j})\right)^2}{\sum_j V_j} \tag{9}$$

follows a χ^2_1 distribution, where O_{1j} and E_{1j} are observed and expected events in group 1 at time t_j, and V_j is the variance.

B.4 Positioning our work

Table 6 situates `STRAND` within the existing landscape of methods for analysing persistence diagrams. Vectorisation methods (top block) produce fixed-dimensional representations suitable for downstream machine learning but provide no native hypothesis test. Statistical-inference methods (middle block) return p-values via permutation or kernel procedures but provide no effect-size summary and no localisation of where, along the persistence axis, two cohorts differ. `STRAND` is the only entry that ticks all four columns: a vectorisation with an integrated stability bound, a hypothesis test, an interpretable effect size, and per-scale localisation through the Kaplan-Meier curve. The closest comparator is the Betti-permutation test of Islambekov and Pathirana [2024], which combines a vectorisation with a permutation test on the same object; the differentiator for `STRAND` is the effect-size and per-scale-localisation machinery that the survival representation makes available natively.

C Proofs

C.1 Proofs of Theorem 3.1 and Propositions 3.2–3.3

We provide a complete proof of the stability theorem for `STRAND` vectorisation.

Table 6: Comparison of methods for analyzing persistence diagrams. STRAND is the only method that simultaneously supports vectorisation for downstream ML, hypothesis testing, and interpretable effect sizes. Methods above the line are vectorisations; methods below the line are statistical tests.

Method	Representation	Stability	Vec.	Test	Eff. Size	Per-scale
Betti Curves [Giusti et al., 2015]	Feature counts	1-Wass. (int.)[¶]	✓	✗	✗	✗
Pers. Images [Adams et al., 2017]	Gauss.-smoothed grid	1-Wass.	✓	✗	✗	✗
Pers. Landscapes [Bubenik and Dłotko, 2017]	Piecewise-linear fn.	∞-Wass.	✓	Ltd.[†]	✗	✗
Pers. Kernels [Reininghaus et al., 2015]	RKHS embedding	1-Wass.	✓	✗[§]	✗	✗
PersLay [Carrière et al., 2020]	Learned aggregation	Cond.[‡]	✓	✗	✗	✗
GRIL [Xin et al., 2023]	Gen. rank invariant	1-Lip.	✓	✗	✗	✗
Wass. permutation [Robinson and Turner, 2017]	Pairwise distances	—	✗	✓	✗	✗
Betti permutation [Islambekov and Pathirana, 2024]	Vec. Betti function	1-Wass.	✓	✓	✗	✗
ALPS [Thomas et al., 2023]	Log lifetime count	—	✗	✓[*]	✗	✗
STRAND **(Ours)**	Pers. survival fn.	1-Wass. (int.)[¶]	✓	✓	✓	✓

[†]Provides z-tests and permutation tests on landscape norms, but without effect size interpretation.
[§]Kernels admit a two-sample test only through an external MMD wrapper, with no native effect size.
[‡]1-Wasserstein stability with specific weight functions; 2-Lipschitz w.r.t. filtration parameters.
[¶]Integrated (L^1) bound, Proposition 3.2, no pointwise bound.
[*]Monte Carlo goodness-of-fit test of a single diagram against a noise model.

C.1.1 Preliminaries and Notation

Definition C.1 (Persistence Diagram)**.** A persistence diagram is a multiset $D = \{(b_i, d_i)\}_{i=1}^n \subset \{(b, d) \in \mathbb{R}^2 : b \le d\}$, together with the diagonal $\Delta = \{(x, x) : x \in \mathbb{R}\}$ with infinite multiplicity.

Definition C.2 (Wasserstein Distance)**.** The 1-Wasserstein distance between persistence diagrams D and D' is:

$$W_1(D, D') = \inf_{\gamma} \sum_{x \in D \cup \Delta} \|x - \gamma(x)\|_\infty, \tag{10}$$

where the infimum is over all bijections $\gamma : D \cup \Delta \to D' \cup \Delta$.

Definition C.3 (Persistence Values)**.** For a persistence diagram $D = \{(b_i, d_i)\}_{i=1}^n$, the persistence values are $p_i = d_i - b_i$ for $i = 1, \dots, n$.

Definition C.4 (Empirical Survival Function)**.** The empirical survival function on persistence values is:

$$\hat{S}_D(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{p_i > t\} = \frac{|\{i : p_i > t\}|}{n}. \tag{11}$$

Definition C.5 (STRAND vectorisation)**.** For grid points $0 \le t_1 < t_2 < \cdots < t_k$, the STRAND survival vectorisation is:

$$\phi_S(D) = \left(\hat{S}_D(t_1), \hat{S}_D(t_2), \dots, \hat{S}_D(t_k)\right) \in [0, 1]^k. \tag{12}$$

C.1.2 Supporting Lemmas

Lemma C.6 (2-Lipschitz Projection)**.** *The persistence projection $\pi : (b, d) \mapsto d - b$ is 2-Lipschitz with respect to the ℓ^∞ norm:*

$$|p_1 - p_2| = |(d_1 - b_1) - (d_2 - b_2)| \le 2 \cdot \|(b_1, d_1) - (b_2, d_2)\|_\infty \tag{13}$$

for any $(b_1, d_1), (b_2, d_2) \in \mathbb{R}^2$.

Proof. We have:

$$\begin{aligned} |(d_1 - b_1) - (d_2 - b_2)| &= |(d_1 - d_2) - (b_1 - b_2)| \\ &\le |d_1 - d_2| + |b_1 - b_2| \quad \text{(triangle inequality)} \\ &\le 2 \max(|d_1 - d_2|, |b_1 - b_2|) \\ &= 2\|(b_1, d_1) - (b_2, d_2)\|_\infty. \end{aligned}$$

□

The constant is attained: for $(b_1, d_1) = (0, 1)$ and $(b_2, d_2) = (\epsilon, 1 - \epsilon)$, the ℓ^∞ distance is ϵ and the persistence values differ by 2ϵ.

Lemma C.7 (Diagonal Matching Cost)**.** *For a point (b, d) with $d > b$ matched to the diagonal, the optimal matching cost is:*

$$\inf_{(c,c)\in\Delta} \|(b, d) - (c, c)\|_\infty = \frac{d - b}{2} = \frac{p}{2}, \tag{14}$$

achieved at $c = \frac{b+d}{2}$. Equivalently, the persistence value satisfies $p = 2 \cdot$ (matching cost). Note: This is a standard result in optimal transport on persistence diagrams; we include the proof for completeness.

Proof. For any $c \in \mathbb{R}$:

$$\|(b, d) - (c, c)\|_\infty = \max(|b - c|, |d - c|). \tag{15}$$

This is minimized when $|b - c| = |d - c|$, which occurs at $c = \frac{b+d}{2}$. The minimum value is:

$$\left| b - \frac{b + d}{2} \right| = \frac{d - b}{2} = \frac{p}{2}. \tag{16}$$

□

Lemma C.8 (Transport Cost on Matched Persistence Values)**.** *Let $D = \{(b_i, d_i)\}_{i=1}^n$ and $D' = \{(b'_j, d'_j)\}_{j=1}^m$ be persistence diagrams, and let γ^* be an optimal matching for $W_1(D, D')$. Define matched persistence values for each $i \in \{1, \ldots, n\}$:*

$$p_i^{(\text{ext})} = \begin{cases} p'_j = d'_j - b'_j & \text{if } (b_i, d_i) \leftrightarrow (b'_j, d'_j) \text{ under } \gamma^* \\ 0 & \text{if } (b_i, d_i) \leftrightarrow \Delta \text{ under } \gamma^* \end{cases} \tag{17}$$

Then:

$$\sum_{i=1}^n |p_i - p_i^{(\text{ext})}| \leq 2 \cdot W_1(D, D'). \tag{18}$$

Proof. We analyze the contribution from each type of match in γ^*.

Case 1: Off-diagonal to off-diagonal. If $(b_i, d_i) \leftrightarrow (b'_j, d'_j)$, then by Lemma C.6:

$$|p_i - p'_j| \leq 2\|(b_i, d_i) - (b'_j, d'_j)\|_\infty = 2 \cdot (\text{matching cost for this pair}). \tag{19}$$

Case 2: Off-diagonal to diagonal. If $(b_i, d_i) \leftrightarrow \Delta$, then $p_i^{(\text{ext})} = 0$ and the contribution is:

$$|p_i - 0| = p_i = 2 \cdot \frac{p_i}{2} = 2 \cdot (\text{matching cost for this pair}), \tag{20}$$

by Lemma C.7.

Summing over all matches:

$$\sum_{i=1}^n |p_i - p_i^{(\text{ext})}| \leq 2 \cdot \sum_{\text{matches involving } D} (\text{matching cost}) \leq 2 \cdot W_1(D, D'). \tag{21}$$

□

C.1.3 Main Proof

Proof of Theorem 3.1. **Step 1: Setup and Matching Partition.**

Let $\gamma^* : D \cup \Delta \to D' \cup \Delta$ be an optimal matching achieving $W_1(D, D')$. Partition the off-diagonal points according to their matching:

- $\mathcal{A} = \{(i, j) : (b_i, d_i) \leftrightarrow (b'_j, d'_j)\}$ — off-diagonal to off-diagonal,
- $\mathcal{B} = \{i : (b_i, d_i) \leftrightarrow \Delta\}$ — points of D matched to diagonal,
- $\mathcal{C} = \{j : \Delta \leftrightarrow (b'_j, d'_j)\}$ — points of D' matched from diagonal.

Let $a = |\mathcal{A}|$, so $|\mathcal{B}| = n - a$ and $|\mathcal{C}| = m - a$.

Note: In practice, when the Wasserstein distance is small relative to persistence values, most off-diagonal points are matched off-diagonal to off-diagonal, with diagonal matching reserved for cardinality differences or very short-lived features.

Step 2: Matched Persistence Values.

For each $i \in \{1, \ldots, n\}$, define matched persistence values as in Lemma C.8:

$$p_i^{(\text{ext})} = \begin{cases} p'_j & \text{if } (b_i, d_i) \leftrightarrow (b'_j, d'_j) \in D' \\ 0 & \text{if } (b_i, d_i) \leftrightarrow \Delta \end{cases} \tag{22}$$

Note: Even when $n = m$, we may have $|\mathcal{B}| = |\mathcal{C}| > 0$. That is, equal cardinality does *not* imply all points are matched off-diagonal to off-diagonal.

Step 3: Decompose the Survival Function Difference.

For any threshold $t \geq 0$, let $N_D(t) = |\{i : p_i > t\}|$ and $N_{D'}(t) = |\{j : p'_j > t\}|$. Then:

$$\hat{S}_D(t) = \frac{N_D(t)}{n}, \qquad \hat{S}_{D'}(t) = \frac{N_{D'}(t)}{m}. \tag{23}$$

We decompose the difference as:

$$\hat{S}_D(t) - \hat{S}_{D'}(t) = \frac{N_D(t)}{n} - \frac{N_{D'}(t)}{m} \tag{24}$$

$$= \underbrace{\left(\frac{N_D(t)}{n} - \frac{N_D(t)}{m}\right)}_{\text{Term (I): normalization difference}} + \underbrace{\left(\frac{N_D(t) - N_{D'}(t)}{m}\right)}_{\text{Term (II): count difference}}. \tag{25}$$

Step 4a: Bound Term (I) — Normalization Difference.

$$\left|\frac{N_D(t)}{n} - \frac{N_D(t)}{m}\right| = N_D(t)\left|\frac{1}{n} - \frac{1}{m}\right| = N_D(t) \cdot \frac{|m-n|}{nm}. \tag{26}$$

Since $N_D(t) \leq n$:

$$|\text{Term (I)}| \leq n \cdot \frac{|m-n|}{nm} = \frac{|m-n|}{m} = \left|1 - \frac{n}{m}\right|. \tag{27}$$

Step 4b: Bound the cardinality difference via diagonal matching.

The bound $|\text{Term (I)}| \leq |1 - n/m|$ from Step 4 can be tightened by observing that the cardinality difference itself is controlled by W_1. Without loss of generality assume $m \geq n$ (the case $n > m$ is symmetric). In any optimal matching γ^* achieving $W_1(D, D')$, the set $\mathcal{C} = \{j : \Delta \leftrightarrow (b'_j, d'_j)\}$ defined in Step 1 satisfies $|\mathcal{C}| = m - a \geq m - n$, since at most $a \leq n$ points of D' can be matched off-diagonal. By Lemma C.7, each diagonally-matched point (b'_j, d'_j) incurs cost $p'_j/2$ in W_1. Therefore:

$$W_1(D, D') \;\geq\; \sum_{j \in \mathcal{C}} \frac{p'_j}{2} \;\geq\; |\mathcal{C}| \cdot \frac{\delta_\Delta}{2} \;\geq\; (m-n) \cdot \frac{\delta_\Delta}{2}, \tag{28}$$

where $\delta_\Delta = \min_{j \in \mathcal{C}} p'_j$. Rearranging:

$$\left|1 - \frac{n}{m}\right| \;=\; \frac{m-n}{m} \;\leq\; \frac{2\,W_1(D, D')}{m \cdot \delta_\Delta} \;=\; \frac{2\,W_1(D, D')}{\max(n, m) \cdot \delta_\Delta}. \tag{29}$$

Step 5: Bound Term (II) — Count Difference via Straddling.

We relate $N_D(t)$ and $N_{D'}(t)$ through the matching γ^*.

Step 5a: Express counts using the matching structure.

$$N_D(t) = \sum_{i=1}^{n} \mathbf{1}\{p_i > t\} \tag{30}$$

$$= \sum_{(i,j)\in\mathcal{A}} \mathbf{1}\{p_i > t\} + \sum_{i\in\mathcal{B}} \mathbf{1}\{p_i > t\}. \tag{31}$$

$$N_{D'}(t) = \sum_{j=1}^{m} \mathbf{1}\{p'_j > t\} \tag{32}$$

$$= \sum_{(i,j)\in\mathcal{A}} \mathbf{1}\{p'_j > t\} + \sum_{j\in\mathcal{C}} \mathbf{1}\{p'_j > t\}. \tag{33}$$

Therefore:

$$N_D(t) - N_{D'}(t) = \sum_{(i,j)\in\mathcal{A}} \big(\mathbf{1}\{p_i > t\} - \mathbf{1}\{p'_j > t\}\big) \tag{34}$$

$$+ \sum_{i\in\mathcal{B}} \mathbf{1}\{p_i > t\} - \sum_{j\in\mathcal{C}} \mathbf{1}\{p'_j > t\}. \tag{35}$$

Step 5b: Unify using matched persistence values.

Recall that for $i \in \mathcal{B}$, we have $p_i^{(\text{ext})} = 0$, so:

$$\mathbf{1}\{p_i > t\} - \mathbf{1}\{p_i^{(\text{ext})} > t\} = \mathbf{1}\{p_i > t\} - \mathbf{1}\{0 > t\} = \mathbf{1}\{p_i > t\} \tag{36}$$

since $\mathbf{1}\{0 > t\} = 0$ for all $t \geq 0$.

Thus:

$$\sum_{i=1}^{n} \Big(\mathbf{1}\{p_i > t\} - \mathbf{1}\{p_i^{(\text{ext})} > t\}\Big) = \sum_{(i,j)\in\mathcal{A}} \big(\mathbf{1}\{p_i > t\} - \mathbf{1}\{p'_j > t\}\big) + \sum_{i\in\mathcal{B}} \mathbf{1}\{p_i > t\}. \tag{37}$$

Step 5c: Define the straddling set.

Definition C.9 (Straddling)**.** A pair $(p_i, p_i^{(\text{ext})})$ *straddles* threshold t if exactly one value exceeds t:

$$\mathcal{S}(t) = \Big\{ i \in \{1, \ldots, n\} : \min(p_i, p_i^{(\text{ext})}) \leq t < \max(p_i, p_i^{(\text{ext})}) \Big\}. \tag{38}$$

We also claim $\Big|\mathbf{1}\{p_i > t\} - \mathbf{1}\{p_i^{(\text{ext})} > t\}\Big| = 1$ if and only if $i \in \mathcal{S}(t)$.

Proof of Claim. The indicator difference equals 1 exactly when one value exceeds t and the other does not:

p_i vs t	$p_i^{(\text{ext})}$ vs t	$\lvert\mathbf{1}\{p_i > t\} - \mathbf{1}\{p_i^{(\text{ext})} > t\}\rvert$	Straddles?
$p_i > t$	$p_i^{(\text{ext})} > t$	0	No
$p_i > t$	$p_i^{(\text{ext})} \leq t$	1	Yes
$p_i \leq t$	$p_i^{(\text{ext})} > t$	1	Yes
$p_i \leq t$	$p_i^{(\text{ext})} \leq t$	0	No

The straddling condition $\min(p_i, p_i^{(\text{ext})}) \leq t < \max(p_i, p_i^{(\text{ext})})$ holds exactly in rows 2 and 3. □

Step 5d: Bound using straddling count.

From Step 5b:

$$\left|\sum_{i=1}^{n}\left(\mathbf{1}\{p_i > t\} - \mathbf{1}\{p_i^{(\mathrm{ext})} > t\}\right)\right| \le \sum_{i=1}^{n}\left|\mathbf{1}\{p_i > t\} - \mathbf{1}\{p_i^{(\mathrm{ext})} > t\}\right| = |\mathcal{S}(t)|. \tag{39}$$

Step 5e: Account for points in $\mathcal{C}$.

The count $N_{D'}(t)$ includes contributions from points in $\mathcal{C}$ (matched from diagonal), which are not captured by the sum over $i \in \{1, \dots, n\}$. We have:

$$N_D(t) - N_{D'}(t) = \sum_{i=1}^{n}\left(\mathbf{1}\{p_i > t\} - \mathbf{1}\{p_i^{(\mathrm{ext})} > t\}\right) - \sum_{j\in\mathcal{C}}\mathbf{1}\{p'_j > t\}. \tag{40}$$

Therefore:

$$|N_D(t) - N_{D'}(t)| \le |\mathcal{S}(t)| + |\mathcal{C}|. \tag{41}$$

Step 5f: Bound $|\mathcal{S}(t)|$ and $|\mathcal{C}|$.

For $i \in \mathcal{S}(t)$, we have $p_i \neq p_i^{(\mathrm{ext})}$. Let $\delta = \min_{i:p_i\neq p_i^{(\mathrm{ext})}} |p_i - p_i^{(\mathrm{ext})}|$. By Lemma C.8, each index with $p_i \neq p_i^{(\mathrm{ext})}$ contributes at least δ to $\sum_i |p_i - p_i^{(\mathrm{ext})}| \le 2W_1(D, D')$, so:

$$|\mathcal{S}(t)| \le \frac{2W_1(D, D')}{\delta}. \tag{42}$$

For $|\mathcal{C}|$: each $j \in \mathcal{C}$ corresponds to $(b'_j, d'_j) \in D'$ matched from the diagonal, contributing $p'_j/2$ to W_1 by Lemma C.7. With $\delta_\Delta = \min_{j\in\mathcal{C}} p'_j$:

$$|\mathcal{C}| \cdot \frac{\delta_\Delta}{2} \le W_1(D, D'), \quad \text{so} \quad |\mathcal{C}| \le \frac{2W_1(D, D')}{\delta_\Delta}. \tag{43}$$

Step 6: Combine the Bounds.

By symmetry of the bound in D and D', assume without loss of generality $m = \max(n, m)$. From Step 4b:

$$|\text{Term (I)}| \le \frac{2W_1(D, D')}{\delta_\Delta \cdot \max(n, m)}. \tag{44}$$

For Term (II), Steps 5e–5f give $|N_D(t) - N_{D'}(t)| \le |\mathcal{S}(t)| + |\mathcal{C}|$. Dividing by $m = \max(n, m)$ and applying Step 5f:

$$|\text{Term (II)}| \le \frac{2W_1(D, D')}{\delta \cdot \max(n, m)} + \frac{2W_1(D, D')}{\delta_\Delta \cdot \max(n, m)}. \tag{45}$$

Combining and using $\max(n, m) \ge \min(n, m)$ to weaken the δ term:

$$|\hat{S}_D(t) - \hat{S}_{D'}(t)| \le |\text{Term (I)}| + |\text{Term (II)}| \tag{46}$$

$$\le \frac{4W_1(D, D')}{\delta_\Delta \cdot \max(n, m)} + \frac{2W_1(D, D')}{\delta \cdot \min(n, m)}. \tag{47}$$

Step 7: Conclude.

Taking the maximum over the grid points $t_1, \dots, t_k$:

$$\|\phi_S(D; \mathbf{t}) - \phi_S(D'; \mathbf{t})\|_\infty \le \frac{4W_1(D, D')}{\delta_\Delta \cdot \max(n, m)} + \frac{2W_1(D, D')}{\delta \cdot \min(n, m)}. \tag{48}$$

If $n = m$ and γ^* matches every point off the diagonal with $p_i = p'_{\gamma^*(i)}$, then $\mathcal{B} = \mathcal{C} = \emptyset$ and $\mathcal{S}(t) = \emptyset$ for every t, so $N_D = N_{D'}$ and the two vectors coincide. The bound holds for every pair of diagrams but does not vanish as $W_1(D, D') \to 0$, since $\delta, \delta_\Delta \le 2\, W_1(D, D')$ (Section 3.1).

□

C.1.4 Proof of Proposition 3.2

Proof. Let γ^* be an optimal matching and $\mathcal{A}, \mathcal{B}, \mathcal{C}$ the partition of Step 1. By Eqs. (34)–(35) and the triangle inequality,

$$\int_0^\infty |N_D(t) - N_{D'}(t)|\, dt \le \sum_{(i,j)\in\mathcal{A}} \int_0^\infty \left|\mathbf{1}\{p_i > t\} - \mathbf{1}\{p'_j > t\}\right| dt + \sum_{i\in\mathcal{B}} \int_0^\infty \mathbf{1}\{p_i > t\}\, dt + \sum_{j\in\mathcal{C}} \int_0^\infty \mathbf{1}\{p'_j > t\}\, dt.$$

For $a, b \ge 0$, $\int_0^\infty |\mathbf{1}\{a > t\} - \mathbf{1}\{b > t\}|\, dt = |a - b|$ and $\int_0^\infty \mathbf{1}\{a > t\}\, dt = a$, so the right-hand side equals $\sum_{\mathcal{A}} |p_i - p'_j| + \sum_{\mathcal{B}} p_i + \sum_{\mathcal{C}} p'_j$. By Lemma C.6, each term of the first sum is at most twice the cost of its pair; by Lemma C.7, each term of the other two sums is exactly twice the cost of matching that point to the diagonal. Summing over the pairs of γ^* gives at most $2\, W_1(D, D')$. When $|D| = |D'| = n$, $\hat{S}_D - \hat{S}_{D'} = (N_D - N_{D'})/n$, which gives Eq. (5). □

Remark C.10 (Unequal cardinality). For $|D| = n$ and $|D'| = m$, writing $\hat{S}_D - \hat{S}_{D'} = N_D\,(1/n - 1/m) + (N_D - N_{D'})/m$ and using $\int_0^\infty N_D(t)\, dt = \sum_i p_i$ gives

$$\int_0^\infty \left|\hat{S}_D(t) - \hat{S}_{D'}(t)\right| dt \;\le\; \left|\frac{1}{n} - \frac{1}{m}\right| \sum_{i=1}^{n} p_i + \frac{2\, W_1(D, D')}{m}.$$

The first term does not vanish with $W_1(D, D')$: it records the change of normalisation when a feature is added or removed.

C.1.5 Proof of Proposition 3.3

Proof. Let γ^* be an optimal matching, so every pair of γ^* costs at most $W_1(D, D') < \eta/2$. A point of D matched to the diagonal would cost $p_i/2 \ge \eta/2$ by Lemma C.7, so $\mathcal{B} = \emptyset$, and since $|D'| = |D|$, also $\mathcal{C} = \emptyset$. For each pair, Lemma C.6 gives $|p_i - p'_{\gamma^*(i)}| \le 2\, W_1(D, D') < \eta \le |p_i - t_l|$ for every grid point t_l, so p_i and $p'_{\gamma^*(i)}$ lie on the same side of every t_l. Hence $N_D(t_l) = N_{D'}(t_l)$ for all l, and since $|D| = |D'|$ the two vectors coincide. □

C.1.6 Remarks on the Stability Bound

Remark C.11 (Data-dependent quantities δ and δ_Δ). Both quantities are at most $2\, W_1(D, D')$ whenever they are defined, which is why Eq. (4) bounds the sup-norm difference without making it small. Propositions 3.2 and 3.3 do not depend on them.

Remark C.12 (Comparison with other methods). Persistence Landscapes satisfy $\|\lambda_D - \lambda_{D'}\|_\infty \le W_\infty(D, D')$ for diagrams of any cardinality [Bubenik and Dłotko, 2017], and Persistence Images satisfy $\|PI(D) - PI(D')\|_\infty \le C(\sigma)\, W_1(D, D')$ with $C(\sigma) \to \infty$ as $\sigma \to 0$ [Adams et al., 2017]. `STRAND`'s grid evaluation has no pointwise guarantee of this kind (Theorem 3.1). Its guarantee is integrated (Proposition 3.2), and the Betti curve satisfies the same integrated bound, $\int |\beta_D - \beta_{D'}| \le 2\, W_1(D, D')$, by the same argument applied to the intervals $[b_i, d_i]$.

D Simulation design and signal–noise discrimination

We first evaluate our approach using survival analysis to distinguish topological signal from noise based on death times across homology groups H_0 (connected components), H_1 (loops), and H_2 (voids). Figure 4 shows results on an illustrative example of a 4D torus with fixed signal probability. We display Kaplan-Meier survival curves on all topological features (A) showing the proportion of features remaining alive at each filtration parameter ϵ. Figure 4 B displays Kaplan-Meier survival curves for each signal-noise features, with associated p-value from the log-rank test. Figure 4 C presents a complementary perspective via death time density distributions. In H_0 and H_1, signal features significantly outlive noise features ($p < 0.0001$, log-rank test). For voids in H_2, we see that signal voids die at the most advanced filtration.

Strong discrimination across signal regimes. Figure 4A shows pronounced discrimination between signal and noise populations across all three homological dimensions. All H_0, H_1, and H_2 exhibit consistently strong discrimination with log-rank p-values below the significance threshold ($p < 0.05$) throughout the signal probability cases, in particular with balanced signal-noise ratios ($0.40 - 0.60$ signal probability).

Effect size patterns reveal survival differences. Figure 4B quantifies the magnitude of survival differences through median survival differences (signal minus noise). All three dimensions exhibit decreasing patterns as signal probability increases. In H_0 and H_1, the transition from positive to negative differences signal probability (around 0.10 and 0.20, respectively) indicates that in low signal cases, noise features die earlier than signal features, while in high signal cases, the reverse occurs. When signal dominates, signal features may merge/die earlier due to increased connectivity. H_2 shows predominantly positive differences, indicating that signal features consistently die at a higher filtration compared to noise features. Notably, median differences approaching zero align with non-significant p-values, confirming the relationship between effect size and statistical significance.

Whether topological features from genuine geometric structures die at systematically different filtration scales compared to noise features, these illustrative results provide affirmative evidence across all homological dimensions (H_0, H_1, and H_2). Signal and noise features exhibit fundamentally different survival characteristics.

Following the simulation settings from Stolz [2023], we evaluate `STRAND` on synthetic point clouds with controlled signal-noise ratios.

Signal. We generate synthetic point clouds containing mixtures of signal points (lying on known topological structures) and noise points. For each dataset, we control the signal probability $p_{signal} \in [0, 1]$, which determines the proportion of points that belong to the underlying topological structure versus noise. **3D Sphere**: Signal points are uniformly distributed on the unit sphere $S^2 \subset \mathbb{R}^3$, generated by normalizing random vectors from a 3D Gaussian distribution. This provides a canonical 2-dimensional manifold with known homology (single connected component, no 1-loop, single void). **4D Torus**: Signal points lie on a 2-torus embedded in $\mathbb{R}^4$ using the parametrization: $(\cos(\gamma), \sin(\gamma), \cos(\phi), \sin(\phi))$, where $\gamma, \phi \in (0, 2\pi)$ are uniformly distributed angles. This flat torus embedding creates a more complex test case with characteristic topological signature (single connected component, two independent 1-loops, single void). **4D Klein Bottle**: Signal points are generated using the Klein bottle parametrization:

$$
\begin{aligned}
x_1 &= \cos(\gamma)(r\cos(\phi) + C) \\
x_2 &= \sin(\gamma)(r\cos(\phi) + C) \\
x_3 &= \cos(\gamma/2) \cdot r\sin(\phi) \\
x_4 &= \sin(\gamma/2) \cdot r\sin(\phi)
\end{aligned}
$$

with $r = 3$, $C = 2$, and $\gamma, \phi \in (0, 2\pi)$. The Klein bottle represents a non-orientable surface with known homology (single connected component, single 1-loop, no voids), testing method robustness on non-orientable manifolds.

Noise. To assess robustness across different noise characteristics, we implement four distinct noise models for the sphere experiments, each probing different aspects of the signal-noise discrimination problem: Cube noise points are uniformly distributed in the cube $[-1, 1]^3$, creating ambient noise that completely surrounds the sphere signal. Plane noise points are confined to the xy-plane, uniformly distributed in $[-3, 3]^2 \times 0$, creating structured noise in a lower-dimensional subspace that intersects the sphere. Line noise points lie on the x-axis with coordinates $(\alpha, 0, 0)$ where α is uniformly distributed in $[-50, 50]$, representing highly structured 1D noise geometrically distant from the 2D sphere signal. Laplace noise points follow the same line structure but with α drawn from a Laplace distribution Laplace($\mu = 4, \sigma = 0.5$), clipped to $[-50, 50]$, testing robustness to different probability distributions within the same geometric structure. For the torus experiments, noise points follow the same parametric form as signal points but with radius r uniformly distributed in $(0, 2)$ instead of being fixed at 1, creating manifold-respecting noise that occupies a different region of the embedding space. For Klein bottle experiments, noise points use the same parametrization but with $r \in [2, 4]$ and $C \in [1, 3]$ uniformly distributed, creating noise that maintains the topological character while being geometrically distinguishable from the signal.

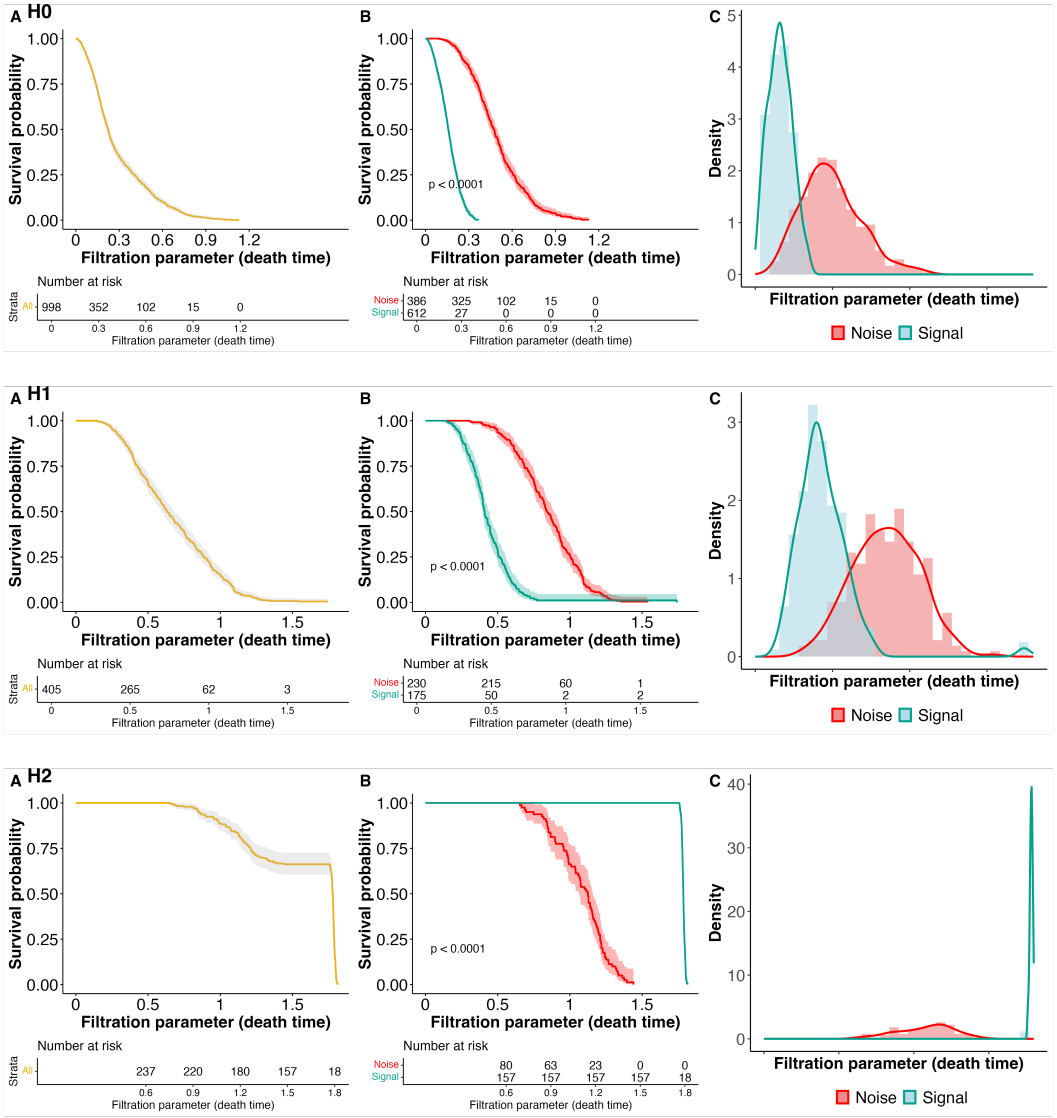


Figure 4: Kaplan-Meier estimators of topological features in 4D torus data with signal parameter of 0.6. Each row corresponds to a different homology dimension computed by persistent homology (e.g., dimension 0 for connected components, 1 for loops, and 2 for voids). (A) Overall survival curves with confidence intervals and risk set tables. (B) Stratified survival curves comparing signal vs. noise components, with p-values from log-rank tests. (C) Density distributions of filtration parameters (death times) for both signal and noise components.

Generation. For each combination of topological object, noise type, and signal probability, we generate 10 independent datasets, each containing $n = 1000$ points. We sample signal probabilities at 50 equally spaced points from 0 to 1, providing detailed coverage of the transition from pure noise ($p_{signal} = 0$) to pure signal ($p_{signal} = 1$). This results in 500 datasets per topological object (50 probability values × 10 simulations each), totaling 2,000 datasets across the four noise conditions for spheres, plus 500 datasets each for torus and Klein bottle experiments. We use different random seeds for each simulation to ensure statistical independence while maintaining reproducibility.

D.1 Single-diagram contrasts

Table 7 reports the submitted calibration, in which each group is a single diagram and the features of that diagram are the sampling units; the classical variance $\sum_j V_j$ is the appropriate one in this design. Figures 2 and 5 are contrasts of the same kind, between the signal and the noise features of one point cloud.

Table 7: Type I error of the classical test for single-diagram contrasts (one diagram per group, 200 replications per setting, Wilson 95% CIs).

Setting	Dim	$\hat{\alpha}$ [95% CI]
Torus	H_0	0.025 [0.011, 0.057]
	H_1	0.075 [0.046, 0.120]
Klein bottle	H_0	0.000 [0.000, 0.019]
	H_1	0.015 [0.005, 0.043]
Sphere	H_0	0.025 [0.011, 0.057]
	H_1	0.025 [0.011, 0.057]

This appendix presents complementary results across five additional topological scenarios examined in our study, demonstrating the generalizability of `STRAND` beyond the 4D torus results presented in the main text. We explore Klein bottle and spheres with four types of noise: line, Laplace, cube, and plane configurations.

D.2 Death scale discrimination

Additional results across all topological scenarios (Figure 5) demonstrate consistency in `STRAND`'s discriminative power. The key patterns mirror those observed in the 4D torus: strong discrimination between signal and noise populations with log-rank p-values consistently below 0.05 across the signal probability spectrum, and characteristic decreasing median survival differences that transition from positive (signal features survive longer) at low signal probabilities to negative at high signal probabilities.

Table 8: Type I error for parametric distributions ($n = 500$ replications).

Distribution	Observed $\hat{\alpha}$	95% CI
Exponential($\lambda = 1$)	0.062	[0.044, 0.087]
Weibull($k = 2$)	0.062	[0.044, 0.087]
LogNormal($\mu = 0, \sigma = 1$)	0.044	[0.029, 0.066]

Interpretation. Signal-vs-signal tests show conservative Type I error because pure manifold samples produce highly reproducible topological structure. This is expected behavior: when both groups are drawn from the same well-defined manifold, their persistence diagrams are nearly identical, making false rejections rare.

E Permutation calibration of the asymptotic log-rank test

Features of the same diagram are dependent: in H_0 they are nested by the elder rule, and in H_k for $k \geq 1$ they are coupled through homological constraints on births and deaths. With $\bar{\rho}$ the average within-diagram correlation and $\bar{n}$ the mean diagram size, the variance of the log-rank numerator

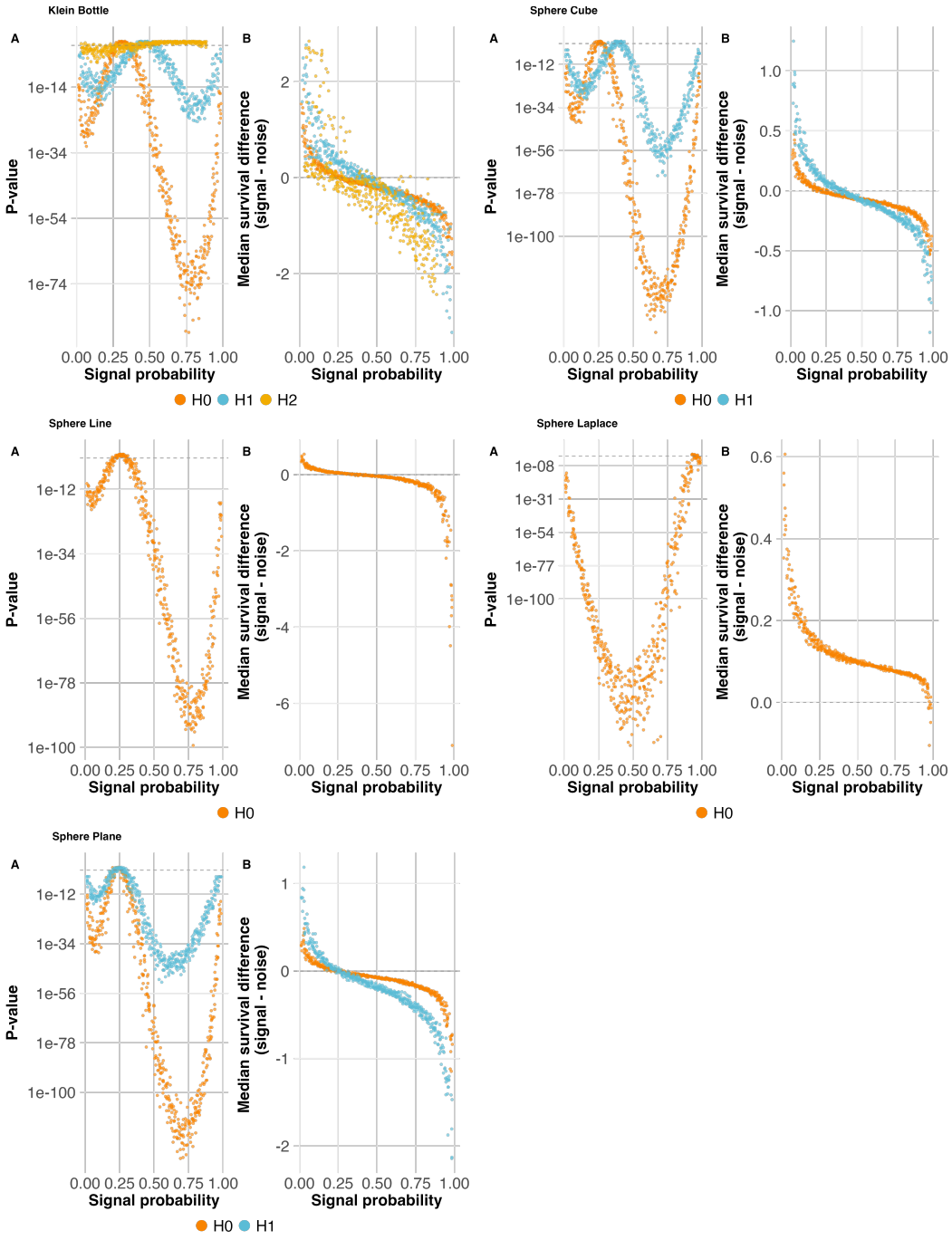


Figure 5: Signal vs. noise discrimination analysis across Klein bottle, and spheres with four type of noise (line, Laplace, cube, plane). Each point represents one simulation from 500 synthetic datasets spanning 50 signal probabilities in [0,1] with 10 independent realizations each. (A) shows log-rank test p-values comparing signal and noise survival distributions, and (B) shows median survival time differences (signal minus noise).

departs from its working-independence form $\sum_j V_j$ by a factor of roughly $1+(\bar{n}-1)\bar{\rho}$. The classical test is anti-conservative when $\bar{\rho} > 0$ and conservative when $\bar{\rho} < 0$, and the sign is not fixed by the elder rule alone. The clustered variance of Eq. (7) removes the dependence on $\bar{\rho}$. In this appendix, we check both tests against a reference that makes no assumption about within-diagram dependence.

Setup. For each of the six manifold scenarios of Table 1 (4D torus, 4D Klein bottle, 3D sphere with cube, plane, line and Laplace noise), in H_0 and H_1, giving 12 conditions, we draw two groups of eight diagrams from the same signal mixture (200 points per cloud, signal probability uniform on $[0, 1]$), 1,000 replications per condition. For each replicate we compute:

- the classical log-rank p-value p_{LR}, with the variance $\sum_j V_j$ of Eq. (6);
- the clustered log-rank p-value p_{cl} of Eq. (7);
- a permutation p-value p_{perm} obtained by reassigning group labels to whole diagrams and recomputing the clustered log-rank statistic, with $B = 1{,}000$ permutations.

Diagram-level permutation preserves within-diagram dependence exactly — whatever correlation exists between persistence values within a diagram is carried through every permuted replicate — and is therefore exchangeable under the null without any independence assumption.

Findings. Table 9 reports, per condition, the agreement of each asymptotic test with the exact diagram-level permutation test. For the clustered test, decision concordance lies between 0.970 and 0.998 and Spearman ρ is at least 0.988. For the classical test, concordance falls to 0.381 on torus H_0 and below 0.65 on every sphere H_0 condition, the conditions with the largest feature counts, which is the anti-conservative direction predicted by positive within-diagram correlation in heterogeneous diagrams; the rank correlation stays high because the classical statistic orders replicates correctly and only misstates their scale.

Table 9: Per-condition agreement of the clustered and of the classical log-rank test with the diagram-level permutation test, in the design of Table 1 (1,000 replications per condition). Concordance: fraction of replicates in agreement on reject/retain at $\alpha = 0.05$; Spearman ρ: rank correlation between the two p-values across replicates.

		Clustered		Classical	
Manifold	Dim	Concordance	Spearman ρ	Concordance	Spearman ρ
Torus (4D)	H_0	0.998	1.000	0.381	0.981
	H_1	0.985	1.000	0.832	0.985
Klein bottle (4D)	H_0	0.990	1.000	0.783	0.986
	H_1	0.992	1.000	0.968	0.978
Sphere, cube noise	H_0	0.997	0.999	0.547	0.975
	H_1	0.990	1.000	0.968	0.978
Sphere, plane noise	H_0	0.978	0.993	0.613	0.927
	H_1	0.998	1.000	0.944	0.977
Sphere, line noise	H_0	0.970	0.988	0.634	0.893
	H_1	0.986	0.999	0.945	0.971
Sphere, Laplace noise	H_0	0.980	0.996	0.501	0.938
	H_1	0.981	0.999	0.934	0.974
Range (12 conditions)		0.970–0.998	0.988–1.000	0.381–0.968	0.893–0.986

Conclusion. The clustered log-rank test is the default in `STRAND`. The diagram-level permutation test is exact under exchangeability of diagrams and is the recommended test when there are few diagrams per group; with equal group sizes it cannot reach $p < 0.05$ below four diagrams per group.

F Power Analysis: Additional Manifolds

We analyse the power of the two-sample test for two groups of ten diagrams on the 4D torus, with signal probability 0.5 in group A and p_B in group B (Table 10). Figure 6 shows empirical power as a function of the achieved hazard ratio for H_0 (connected components) and H_1 (loops) features.

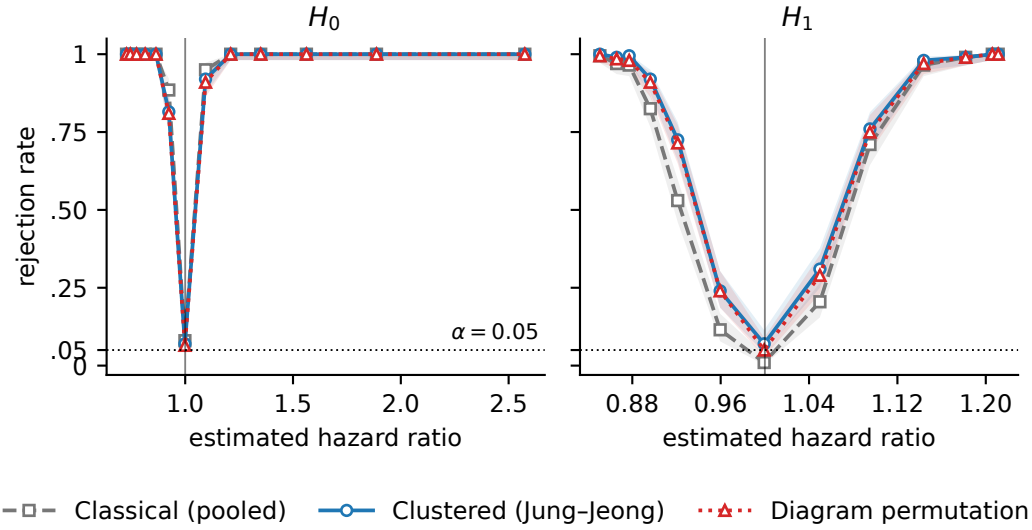


Figure 6: Power vs hazard ratio for `STRAND` two-sample test on simulated torus data. Left: H_0 features. Right: H_1 features. The rejection rate is close to $\alpha = 0.05$ at $\mathrm{HR} = 1$ (no true difference) and reaches 1 as $|\log(\mathrm{HR})|$ increases.

Table 10: Empirical rejection rate at $\alpha = 0.05$ as a function of the hazard ratio (4D torus, 10 diagrams per group, 200 replications, Wilson 95% CIs). Group A has signal probability 0.5; group B has p_B, and the hazard ratio is the mean achieved ratio, which this design can move between 0.73 and 1.89 at H_0 and between 0.85 and 1.21 at H_1. Classical: variance $\sum_j V_j$; Clustered: Eq. (7); Permutation: labels reassigned to whole diagrams, $B = 1{,}000$.

Dim	p_B	Achieved HR	Classical	Clustered [95% CI]	Permutation [95% CI]
H_0	0.05	0.728	1.000	1.000 [0.981, 1.000]	1.000 [0.981, 1.000]
	0.5	0.999	0.080	0.070 [0.042, 0.114]	0.065 [0.038, 0.108]
	0.8	1.562	1.000	1.000 [0.981, 1.000]	1.000 [0.981, 1.000]
	0.875	1.887	1.000	1.000 [0.981, 1.000]	1.000 [0.981, 1.000]
H_1	0.95	0.851	0.995	1.000 [0.981, 1.000]	0.995 [0.972, 0.999]
	0.5	0.999	0.010	0.070 [0.042, 0.114]	0.050 [0.027, 0.090]
	0.125	1.211	1.000	1.000 [0.981, 1.000]	1.000 [0.981, 1.000]

Key findings include: (1) power reaches 1 as soon as the achieved HR departs from 1 by 0.15 or more, and (2) at $\mathrm{HR} = 1$ the clustered test rejects 0.070 of the time in both dimensions, whereas the classical test is anti-conservative at H_0 (0.080) and conservative at H_1 (0.010).

F.1 Power against PD-inference baselines

At the signal contrast of Section 4.1 every test reaches full power from five diagrams per group, so Table 11 uses a weaker contrast ($p_{\mathrm{signal}} = 0.6$ against 0.4). At H_0 all four tests reach full power. At H_1 the two tests that compare full diagrams, R&T and MMD-SWK, are more powerful than `STRAND` at $n = 5$ (0.975 and 0.945 against 0.615) and reach full power at $n = 10$, where `STRAND` reaches 0.935. The difference is the information the lifetime marginal discards (Appendix I.7), and `STRAND` returns in exchange an effect size, a confidence interval and a localisation along the persistence axis that neither comparator provides.

Table 11: Power at $\alpha = 0.05$ on simulated 4D torus data at signal probabilities 0.6 against 0.4, two groups of n diagrams, 200 replications; permutation tests with $B = 1{,}000$.

Dim	n	STRAND (clustered)	STRAND (permutation)	R&T	MMD-SWK
H_0	5	1.000	1.000	1.000	1.000
	10	1.000	1.000	1.000	1.000
H_1	5	0.615	0.555	0.975	0.945
	10	0.935	0.925	1.000	1.000

Table 12: Dataset statistics. Molecular datasets (top) represent chemical compounds; social network datasets (middle) represent user interactions; 3D point cloud datasets (bottom) are sampled from surface meshes.

Dataset	Samples	Classes	Domain
MUTAG	188	2	Mutagenicity
PTC-MR	344	2	Toxicity
PROTEINS	1,113	2	Protein structure
ENZYMES	600	6	Enzyme function
NCI1	4,110	2	Anti-cancer activity
DHFR	467	2	Drug activity
COX2	467	2	Drug activity
DD	1,178	2	Protein structure
IMDB-B	1,000	2	Movie collaboration
IMDB-M	1,500	3	Movie collaboration
SwissCheese	1,000	2	Synthetic 3D point clouds
Corals	83	2	3D coral surface meshes
MCB	2,441	21	CAD mechanical components
ModelNet10	4,899	10	CAD object meshes

G Benchmark

G.1 Datasets

Table 12 summarizes dataset characteristics. This benchmark selection follows established practice in topological machine learning Carriere et al. [2017], Adams et al. [2017] and provides diversity across: (i) graph sizes (14–284 average nodes), (ii) classification difficulty (binary and multi-class), (iii) domain semantics (molecular structure vs. social dynamics), and (iv) class balance characteristics. The molecular datasets test whether STRAND captures chemically meaningful topological features, while social network datasets assess generalization to domains where filtration semantics differ.

G.2 Hyperparameter selection

Range normalization. When comparing PDs from different data instances, filtration ranges may vary. We adopt local normalization, i.e. each diagram's grid spans its own $[t_{\min}, t_{\max}]$. This preserves the relative shape of survival curves while allowing comparison across diagrams with different absolute scales. Global normalization (a fixed range across all diagrams) is an alternative when the full dataset is available at vectorisation time.

For all baselines, hyperparameters were selected by grid search under the common cross-validation protocol of Section 4.2 (stratified 10-fold for graphs; the 10 predefined splits of Graf et al. [2026] for point clouds), with selection performed on the training portion of each split only. Table 13 lists the search grids.

G.3 Results

Tables 14–21 present complete results for all STRAND variants and baselines across all metrics and classifiers. Values are reported as mean $\pm$ standard deviation over 10-fold cross-validation.

Table 13: Hyperparameter search grids.

Method / Component	Search grid
Persistence kernels	
SWK	bandwidth $\in \{0.1, 1, 10\}$; # directions $= 10$
PSSK	bandwidth $\sigma \in \{0.1, 1, 10\}$
Classifiers	
SVM (RBF, vectorisations)	$C \in \{10^{-2}, 10^{-1}, 1, 10, 10^{2}\}$; $\gamma \in \{\text{scale}, 10^{-3}, 10^{-2}, 10^{-1}\}$
SVM (precomputed, kernels)	$C \in \{10^{-2}, 10^{-1}, 1, 10, 10^{2}\}$
Random Forest	# trees $\in \{100, 300, 500\}$; max depth $\in \{\text{None}, 10, 20\}$; max features $\in \{\sqrt{\cdot}, \log_2\}$

Table 14: Accuracy (%) for all methods (SVM). S=STRAND, B=STRAND-B (birth entry), F=STRAND-F (birth features). Numbers indicate grid size. Bold indicates best per row.

Dataset	S25	S50	S100	B25	B50	B100	F25	F50	F100	PI	PL	BC
COX2	78.2±0.8	78.2±0.8	78.2±0.8	78.2±0.8	78.2±0.8	78.2±0.8	78.5±1.1	78.5±1.1	78.5±1.1	78.7±1.2	78.2±0.8	**78.8**±1.4
DD	70.4±2.5	70.0±2.7	69.6±2.5	71.3±2.8	69.9±2.6	69.4±2.5	71.1±2.8	70.9±2.5	70.3±3.1	75.0±3.2	72.1±3.5	**75.4**±3.8
DHFR	64.7±4.4	64.1±4.6	62.8±4.5	60.8±1.2	60.8±0.6	61.0±0.5	**67.6**±4.4	66.5±5.0	66.8±4.9	64.3±4.6	61.0±0.5	62.6±3.2
ENZYMES	25.6±4.9	24.9±4.6	24.1±4.0	22.6±4.2	21.7±4.4	21.9±4.1	**27.0**±4.2	26.8±3.4	26.4±3.8	24.3±4.2	25.2±4.9	22.0±3.7
IMDB-BINARY	63.6±3.7	63.6±3.7	64.0±3.8	61.2±4.8	61.1±4.8	60.8±5.1	63.7±4.2	63.8±4.1	**64.2**±4.2	62.5±4.0	63.7±4.0	63.4±3.8
IMDB-MULTI	41.0±3.2	**41.3**±3.5	41.2±2.9	39.6±2.9	39.6±2.9	39.6±2.9	40.5±3.6	40.7±3.2	40.5±3.2	40.3±4.0	39.2±2.7	38.4±2.8
MUTAG	77.9±8.7	74.9±9.1	75.8±7.7	71.3±8.6	69.0±5.7	67.2±3.8	74.9±7.9	75.8±7.7	74.0±6.9	**78.8**±8.2	72.2±7.5	78.4±8.3
NCI1	62.3±2.0	62.2±2.0	62.3±1.8	62.4±2.0	62.6±2.1	62.4±2.1	62.5±1.9	62.4±2.0	61.9±1.9	**63.0**±2.0	60.4±2.5	62.0±1.8
PROTEINS	66.4±3.6	65.7±3.8	65.8±3.7	65.8±3.2	65.7±3.1	65.7±3.1	66.4±3.5	66.1±3.5	66.2±3.2	**71.0**±4.2	70.6±4.4	70.0±4.5
PTC-MR	56.8±5.5	56.9±6.0	55.9±5.4	**58.5**±2.8	57.6±2.4	57.6±2.4	57.0±5.6	56.0±5.2	56.0±5.2	55.3±6.5	54.7±8.0	53.6±6.0
Avg	60.7	60.2	60.0	59.2	58.6	58.4	60.9	60.7	60.5	**61.3**	59.7	60.5

PI keeps its accuracy at lower resolution, with 61.6% and 61.7% at 10×10 and 25×25 (200 and 1,250 features) in a single 10-fold run, so STRAND still uses at least $3\times$ fewer features than PI at its accuracy.

Table 15: Accuracy (%) for all methods (RF). S=STRAND, B=STRAND-B (birth entry), F=STRAND-F (birth features). Numbers indicate grid size. Bold indicates best per row.

Dataset	S25	S50	S100	B25	B50	B100	F25	F50	F100	PI	PL	BC
COX2	76.8±3.8	77.0±4.2	76.5±4.1	76.2±3.1	76.2±3.1	75.9±3.0	77.5±3.7	76.9±3.6	77.1±3.7	77.5±3.9	**78.2**±0.8	77.9±4.2
DD	73.3±2.4	73.1±2.7	71.8±3.1	73.3±3.3	73.1±3.5	72.8±3.3	74.1±2.4	73.1±3.0	73.1±3.3	**75.2**±3.5	72.3±3.2	74.9±3.7
DHFR	71.0±4.0	70.9±4.3	71.1±4.0	66.6±5.0	66.2±4.7	66.0±4.6	72.0±4.0	72.0±3.9	72.1±4.0	**72.5**±4.3	61.0±0.5	71.3±4.2
ENZYMES	23.4±5.4	22.4±4.6	21.8±4.9	23.8±4.4	23.9±4.6	23.1±4.4	27.6±4.4	27.6±4.6	27.7±5.1	27.3±4.8	**27.8**±6.0	27.6±5.4
IMDB-BINARY	62.5±4.1	63.8±3.5	62.8±3.0	64.7±3.4	**65.3**±3.7	65.0±3.0	64.9±4.4	65.1±4.4	64.9±4.2	64.1±4.8	61.7±5.0	61.4±4.1
IMDB-MULTI	38.9±3.1	40.2±3.1	**41.8**±3.3	38.0±2.8	38.3±2.9	38.0±2.6	41.3±3.0	41.2±3.0	41.4±2.9	41.2±3.4	41.5±2.8	39.9±2.7
MUTAG	77.7±8.5	77.7±8.5	77.7±8.5	70.8±9.6	70.8±9.6	70.8±9.6	77.7±8.5	77.7±8.5	77.7±8.5	**77.9**±8.4	73.8±7.7	76.6±8.6
NCI1	61.7±2.1	61.5±2.1	61.4±2.2	62.6±2.3	62.6±2.4	62.6±2.3	62.1±2.1	62.1±2.1	61.8±1.9	**63.1**±2.1	61.0±2.5	62.0±1.9
PROTEINS	68.3±5.1	67.9±5.1	67.6±5.0	67.5±4.3	67.4±4.5	67.3±4.6	68.8±4.4	68.7±4.5	68.4±4.3	**70.7**±4.0	68.9±3.8	70.4±4.4
PTC-MR	51.6±6.7	51.6±6.7	51.6±6.9	**58.1**±7.8	**58.1**±7.8	57.9±7.7	51.7±7.6	51.9±7.3	51.8±7.2	54.2±7.0	54.7±8.0	54.4±7.6
Avg	60.5	60.6	60.4	60.2	60.2	59.9	61.8	61.6	61.6	**62.4**	60.1	61.6

Table 16: Bal. Acc. (%) for all methods (SVM). S=STRAND, B=STRAND-B (birth entry), F=STRAND-F (birth features). Numbers indicate grid size. Bold indicates best per row.

Dataset	S25	S50	S100	B25	B50	B100	F25	F50	F100	PI	PL	BC
COX2	50.0±0.0	50.0±0.0	50.0±0.0	50.0±0.0	50.0±0.0	50.0±0.0	50.8±1.8	50.8±1.8	50.8±1.8	**51.5**±2.5	50.0±0.0	**51.5**±2.6
DD	67.2±2.9	66.9±3.2	66.4±2.9	68.1±3.2	66.4±2.8	65.7±2.8	68.0±3.2	67.8±2.8	66.9±3.4	73.1±3.2	68.9±3.8	**73.3**±3.9
DHFR	60.4±5.1	60.2±5.2	58.9±5.3	50.1±1.2	49.9±0.4	50.0±0.0	**65.5**±4.6	64.5±5.3	65.4±5.0	60.6±5.1	50.0±0.0	54.4±3.4
ENZYMES	25.6±4.9	24.9±4.6	24.1±4.0	22.6±4.2	21.7±4.4	21.9±4.1	**27.0**±4.2	26.8±3.4	26.4±3.8	24.3±4.2	25.2±4.9	22.0±3.7
IMDB-BINARY	63.6±3.7	63.6±3.7	64.0±3.8	61.2±4.8	61.1±4.8	60.8±5.1	63.7±4.2	63.8±4.1	**64.2**±4.2	62.5±4.0	63.7±4.0	63.4±3.8
IMDB-MULTI	41.0±3.2	**41.3**±3.5	41.2±2.9	39.6±2.9	39.6±2.9	39.6±2.9	40.5±3.6	40.7±3.2	40.5±3.2	40.3±4.0	39.2±2.7	38.4±2.8
MUTAG	71.0±10.3	67.9±10.8	67.4±8.9	62.8±9.5	55.3±7.0	51.4±4.8	67.4±9.1	67.4±8.9	63.8±8.6	**75.2**±10.0	62.3±8.4	72.7±10.5
NCI1	62.3±2.0	62.3±2.0	62.3±1.8	62.4±2.0	62.6±2.1	62.4±2.1	62.5±1.9	62.4±1.9	62.0±1.9	**63.1**±2.0	60.4±2.5	62.1±1.7
PROTEINS	60.8±4.1	60.0±4.2	60.0±4.2	59.0±3.7	58.9±3.7	58.9±3.7	60.4±4.1	60.0±4.0	60.1±3.8	**68.5**±4.4	67.4±4.4	67.1±4.8
PTC-MR	53.7±5.8	53.7±6.3	52.5±5.4	53.1±2.8	52.0±2.4	52.0±2.4	**53.9**±5.9	52.8±5.4	52.8±5.4	52.4±5.9	52.6±7.5	51.0±5.9
Avg	55.6	55.1	54.7	52.9	51.8	51.3	56.0	55.7	55.3	**57.1**	54.0	55.6

Table 17: Bal. Acc. (%) for all methods (RF). S=STRAND, B=STRAND-B (birth entry), F=STRAND-F (birth features). Numbers indicate grid size. Bold indicates best per row.

Dataset	S25	S50	S100	B25	B50	B100	F25	F50	F100	PI	PL	BC
COX2	56.4±5.4	56.8±6.2	56.5±6.1	50.9±3.5	50.9±3.3	50.3±3.0	57.3±5.5	56.6±5.1	57.2±5.6	57.5±5.3	50.0±0.0	**57.5**±6.5
DD	71.4±2.8	71.3±3.0	69.8±3.5	71.4±3.5	71.0±3.7	70.8±3.5	72.3±2.5	71.1±3.3	71.1±3.5	**73.8**±3.6	70.0±3.5	72.9±3.9
DHFR	68.8±3.9	68.6±4.2	69.0±3.9	64.0±5.2	63.7±4.9	63.5±4.7	70.1±4.0	70.0±3.8	70.1±3.7	**70.4**±4.5	50.0±0.0	69.2±4.5
ENZYMES	23.4±5.4	22.4±4.6	21.8±4.9	23.8±4.4	23.9±4.6	23.1±4.4	27.6±4.4	27.6±4.6	27.7±5.1	27.3±4.8	**27.8**±6.0	27.6±5.4
IMDB-BINARY	62.5±4.1	63.8±3.5	62.8±3.0	64.7±3.4	**65.3**±3.7	65.0±3.0	64.9±4.4	65.1±4.4	64.9±4.2	64.1±4.8	61.7±5.0	61.4±4.1
IMDB-MULTI	38.9±3.1	40.2±3.1	**41.8**±3.3	38.0±2.8	38.3±2.9	38.0±2.6	41.3±3.0	41.2±3.0	41.4±2.9	41.2±3.4	41.5±2.8	39.9±2.7
MUTAG	74.9±9.5	74.9±9.5	**75.0**±9.5	63.6±10.3	63.6±10.3	63.6±10.3	74.9±9.5	74.9±9.5	74.9±9.5	74.8±9.7	65.2±8.9	73.6±9.6
NCI1	61.7±2.1	61.6±2.1	61.4±2.2	62.6±2.3	62.6±2.4	62.6±2.3	62.1±2.1	62.1±2.1	61.8±1.9	**63.1**±2.1	61.0±2.5	62.0±1.9
PROTEINS	66.1±5.2	65.7±5.1	65.4±5.1	65.3±4.2	65.3±4.4	65.2±4.6	66.7±4.6	66.6±4.7	66.4±4.6	**68.7**±4.0	66.9±3.9	68.5±4.6
PTC-MR	50.7±6.4	50.6±6.4	50.7±6.6	**56.4**±7.8	56.4±7.7	56.1±7.6	50.7±7.2	50.9±7.0	50.8±6.9	53.5±6.7	53.6±7.9	53.8±7.1
Avg	57.5	57.6	57.4	56.1	56.1	55.8	58.8	58.6	58.6	**59.4**	54.8	58.6

Table 18: F1 (%) for all methods (SVM). S=STRAND, B=STRAND-B (birth entry), F=STRAND-F (birth features). Numbers indicate grid size. Bold indicates best per row.

Dataset	S25	S50	S100	B25	B50	B100	F25	F50	F100	PI	PL	BC
COX2	68.6±1.1	68.6±1.1	68.6±1.1	68.6±1.1	68.6±1.1	68.6±1.1	69.4±2.1	69.4±2.1	69.4±2.1	**70.0**±2.4	68.6±1.1	70.0±2.7
DD	69.1±3.0	68.7±3.3	68.3±3.0	70.0±3.2	68.3±2.9	67.6±2.9	69.9±3.2	69.7±2.8	68.8±3.5	74.6±3.2	70.8±3.8	**74.9**±3.9
DHFR	63.0±4.9	62.7±5.0	61.4±5.0	46.9±1.9	46.1±0.5	46.2±0.6	**67.3**±4.4	66.2±5.1	66.7±4.9	63.0±5.1	46.2±0.6	55.8±4.0
ENZYMES	24.7±5.4	23.5±5.2	22.5±4.7	21.5±4.2	20.4±4.6	20.7±4.1	**26.6**±4.5	26.3±3.7	26.0±4.0	23.0±4.5	22.7±5.1	20.1±4.3
IMDB-BINARY	62.6±3.8	62.6±3.8	62.8±4.0	60.9±4.7	60.8±4.7	60.6±5.0	62.9±4.2	62.9±4.1	**63.3**±4.2	61.8±4.0	62.8±4.1	62.5±3.8
IMDB-MULTI	37.3±3.7	**37.8**±4.0	37.7±3.3	31.5±2.4	31.5±2.4	31.5±2.4	31.5±2.4	37.0±4.2	37.5±3.7	37.1±3.6	37.2±4.7	34.4±3.5	32.8±3.8
MUTAG	76.1±10.0	72.9±10.2	73.1±9.0	68.6±9.6	60.1±8.1	54.7±6.6	72.8±8.8	73.1±9.0	69.9±9.0	**78.1**±8.7	67.5±9.3	76.8±9.5
NCI1	61.5±2.1	61.4±2.0	61.5±1.9	61.3±2.1	61.6±2.3	61.1±2.2	61.4±2.1	61.3±2.2	61.1±2.0	**62.1**±2.2	60.0±2.6	61.4±1.9
PROTEINS	62.7±4.6	61.7±4.7	61.7±4.7	59.9±4.5	59.9±4.5	59.9±4.5	62.0±4.7	61.5±4.7	61.6±4.4	**70.4**±4.3	69.6±4.4	69.1±4.7
PTC-MR	52.3±7.2	52.2±8.0	50.7±7.0	45.6±5.3	43.6±4.6	43.6±4.6	**52.6**±7.5	51.3±7.1	51.2±7.0	51.3±6.1	52.5±7.4	50.4±6.4
Avg	57.8	57.2	56.8	53.5	52.1	51.4	58.2	57.9	57.5	**59.1**	55.5	57.4

Table 19: F1 (%) for all methods (RF). S=STRAND, B=STRAND-B (birth entry), F=STRAND-F (birth features). Numbers indicate grid size. Bold indicates best per row.

Dataset	S25	S50	S100	B25	B50	B100	F25	F50	F100	PI	PL	BC
COX2	73.2±4.1	73.3±4.7	73.0±4.6	69.5±2.9	69.5±2.9	69.0±2.6	73.9±3.9	73.3±3.7	73.6±4.0	74.0±4.0	68.6±1.1	**74.2**±4.8
DD	72.9±2.6	72.7±2.8	71.3±3.3	72.9±3.4	72.6±3.7	72.3±3.4	73.7±2.4	72.6±3.1	72.7±3.4	**75.0**±3.5	71.7±3.4	74.4±3.8
DHFR	70.7±3.9	70.6±4.2	70.8±3.9	66.1±5.1	65.8±4.8	65.6±4.6	71.8±4.0	71.8±3.8	71.9±3.8	**72.2**±4.3	46.2±0.6	71.0±4.3
ENZYMES	23.1±5.6	22.0±4.6	21.5±4.9	23.3±4.5	23.2±4.6	22.4±4.4	27.1±4.4	27.1±4.7	27.3±5.2	26.8±5.0	**27.3**±6.1	27.1±5.4
IMDB-BINARY	62.3±4.1	63.7±3.6	62.7±3.0	64.0±3.5	64.6±3.9	64.5±3.2	64.8±4.4	**65.0**±4.4	64.7±4.3	64.0±4.8	61.6±5.0	61.3±4.0
IMDB-MULTI	35.6±3.4	36.7±3.4	**38.6**±3.8	34.1±3.2	34.2±3.2	34.0±3.0	37.9±3.6	37.9±3.4	38.1±3.4	37.8±3.9	38.2±3.3	36.2±2.8
MUTAG	77.4±8.4	77.4±8.4	**77.4**±8.4	68.4±10.5	68.4±10.5	68.4±10.5	77.4±8.4	77.4±8.4	77.4±8.4	**77.4**±8.5	70.3±9.1	76.2±8.6
NCI1	61.5±2.0	61.4±2.0	61.2±2.1	62.2±2.4	62.3±2.4	62.3±2.4	61.9±2.0	61.8±2.1	61.5±1.9	**62.9**±2.1	60.6±2.5	61.7±1.9
PROTEINS	67.8±5.2	67.4±5.1	67.1±5.1	67.1±4.2	67.0±4.4	66.9±4.6	68.4±4.4	68.3±4.6	68.0±4.4	**70.3**±3.9	68.5±3.8	70.0±4.4
PTC-MR	50.9±6.5	50.9±6.4	50.9±6.6	**56.6**±8.0	56.6±7.9	56.3±7.8	51.0±7.3	51.1±7.1	51.1±7.0	53.5±6.9	53.8±7.7	53.8±7.5
Avg	59.5	59.6	59.5	58.4	58.4	58.2	60.8	60.6	60.6	**61.4**	56.7	60.6

Table 20: MCC for all methods (SVM). S=STRAND, B=STRAND-B (birth entry), F=STRAND-F (birth features). Numbers indicate grid size. Bold indicates best per row.

Dataset	S25	S50	S100	B25	B50	B100	F25	F50	F100	PI	PL	BC
COX2	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.05±0.10	0.05±0.10	0.05±0.10	0.08±0.13	0.00±0.00	**0.08**±0.13
DD	0.38±0.06	0.37±0.06	0.36±0.06	0.40±0.06	0.37±0.06	0.35±0.06	0.39±0.06	0.39±0.06	0.37±0.07	0.48±0.07	0.41±0.08	**0.49**±0.08
DHFR	0.23±0.10	0.22±0.11	0.19±0.11	0.01±0.08	-0.01±0.03	0.00±0.00	**0.32**±0.09	0.29±0.11	0.31±0.10	0.22±0.11	0.00±0.00	0.13±0.10
ENZYMES	0.11±0.06	0.10±0.06	0.09±0.05	0.07±0.05	0.06±0.05	0.06±0.05	**0.13**±0.05	0.12±0.04	0.12±0.05	0.09±0.05	0.11±0.06	0.07±0.05
IMDB-BINARY	0.29±0.08	0.29±0.08	0.30±0.08	0.23±0.10	0.23±0.10	0.22±0.10	0.29±0.09	0.29±0.09	**0.30**±0.09	0.26±0.08	0.29±0.09	0.28±0.08
IMDB-MULTI	0.13±0.05	**0.13**±0.06	0.13±0.05	0.11±0.05	0.11±0.05	0.11±0.05	0.12±0.06	0.12±0.05	0.12±0.05	0.11±0.07	0.10±0.04	0.09±0.05
MUTAG	0.48±0.22	0.41±0.23	0.43±0.20	0.31±0.23	0.17±0.20	0.04±0.12	0.41±0.20	0.43±0.20	0.36±0.21	**0.53**±0.19	0.33±0.19	0.50±0.20
NCI1	0.26±0.04	0.26±0.04	0.26±0.04	0.26±0.04	0.27±0.04	0.27±0.04	0.26±0.04	0.26±0.04	0.25±0.04	**0.28**±0.04	0.21±0.05	0.25±0.04
PROTEINS	0.27±0.10	0.25±0.10	0.26±0.10	0.26±0.09	0.26±0.09	0.26±0.09	0.27±0.10	0.26±0.09	0.27±0.09	**0.39**±0.09	0.38±0.10	0.36±0.10
PTC-MR	0.09±0.14	0.08±0.15	0.06±0.15	**0.15**±0.12	0.10±0.11	0.10±0.11	0.09±0.15	0.06±0.14	0.06±0.14	0.07±0.14	0.07±0.17	0.03±0.14
Avg	0.22	0.21	0.21	0.18	0.16	0.14	0.23	0.23	0.22	**0.25**	0.19	0.23

Table 21: MCC for all methods (RF). S=STRAND, B=STRAND-B (birth entry), F=STRAND-F (birth features). Numbers indicate grid size. Bold indicates best per row.

Dataset	S25	S50	S100	B25	B50	B100	F25	F50	F100	PI	PL	BC
COX2	0.18±0.15	0.18±0.16	0.17±0.16	0.04±0.15	0.05±0.15	0.02±0.13	0.21±0.14	0.18±0.13	0.19±0.14	**0.21**±0.15	0.00±0.00	0.21±0.17
DD	0.44±0.05	0.44±0.06	0.41±0.07	0.44±0.07	0.44±0.08	0.43±0.07	0.46±0.05	0.44±0.07	0.44±0.07	**0.48**±0.07	0.42±0.07	0.48±0.08
DHFR	0.39±0.08	0.38±0.09	0.39±0.08	0.29±0.11	0.28±0.10	0.28±0.10	0.41±0.08	0.41±0.08	0.41±0.08	**0.42**±0.09	0.00±0.00	0.39±0.09
ENZYMES	0.08±0.07	0.07±0.06	0.06±0.06	0.09±0.05	0.09±0.06	0.08±0.05	0.13±0.05	0.13±0.06	0.13±0.06	0.13±0.06	**0.13**±0.07	0.13±0.06
IMDB-BINARY	0.25±0.08	0.28±0.07	0.26±0.06	0.31±0.07	**0.32**±0.07	0.31±0.06	0.30±0.09	0.30±0.09	0.30±0.09	0.28±0.10	0.24±0.10	0.23±0.08
IMDB-MULTI	0.09±0.05	0.12±0.05	**0.14**±0.05	0.08±0.05	0.08±0.05	0.08±0.04	0.13±0.05	0.13±0.05	0.14±0.05	0.13±0.06	0.14±0.05	0.11±0.05
MUTAG	0.51±0.18	0.51±0.18	0.51±0.18	0.31±0.24	0.31±0.24	0.31±0.24	0.51±0.18	0.51±0.18	0.51±0.18	**0.51**±0.19	0.39±0.19	0.49±0.19
NCI1	0.24±0.04	0.23±0.04	0.23±0.04	0.26±0.05	0.26±0.05	0.26±0.05	0.25±0.04	0.25±0.04	0.24±0.04	**0.27**±0.04	0.22±0.05	0.24±0.04
PROTEINS	0.33±0.11	0.32±0.11	0.32±0.11	0.32±0.09	0.31±0.09	0.31±0.10	0.34±0.09	0.34±0.10	0.34±0.09	**0.39**±0.08	0.35±0.08	0.38±0.09
PTC-MR	0.02±0.13	0.01±0.13	0.02±0.14	0.14±0.17	**0.14**±0.16	0.13±0.16	0.02±0.15	0.02±0.15	0.02±0.14	0.07±0.14	0.08±0.16	0.08±0.15
Avg	0.25	0.25	0.25	0.23	0.23	0.22	0.28	0.27	0.27	**0.29**	0.20	0.27

H Further analysis on ABIDE dataset

Within-subject dependence and the weighted estimand. Table 22 reports, for each cell of Table 5, the design effect (the ratio of the classical variance $\sum_j V_j$ to the clustered variance of Eq. (7)), the intraclass correlation of persistence values within a subject's diagram, and the hazard ratio with and without inverse-size weighting. At H_0 every diagram in an atlas has the same number of features ($m - 1$ for Vietoris–Rips on m ROIs: 399, 199 and 89), so the weights are constant and the two estimators coincide. At H_1 children carry more loops than adults on every atlas (means 369.1 vs 358.4 at Schaefer-400, 166.6 vs 157.7 at CC200, 52.6 vs 49.5 at AAL90), the informative cluster size case of Williamson et al. [2003], and the weighted estimator is the reported one; the two differ by at most 0.003 with no change of decision.

Table 22: Design effect, intraclass correlation and hazard ratios for the ABIDE contrast. HR (weighted) is the value reported in Table 5; 95% CIs from 1,000 subject-bootstrap resamples.

Dim	Atlas	Design effect	ICC	HR (unweighted)	HR (weighted)
H_0	Schaefer-400	76	0.24	0.815 [0.731, 0.909]	0.815 [0.731, 0.909]
	CC200	40	0.21	0.982 [0.881, 1.099]	0.982 [0.881, 1.099]
	AAL90	14.6	0.20	0.967 [0.878, 1.072]	0.967 [0.878, 1.072]
H_1	Schaefer-400	8.1	0.011	1.145 [1.106, 1.185]	1.148 [1.110, 1.191]
	CC200	4.3	0.009	1.101 [1.060, 1.143]	1.103 [1.060, 1.147]
	AAL90	2.0	0.007	1.090 [1.041, 1.142]	1.092 [1.041, 1.148]

Robustness across sites and motion. We assess robustness with a pre-specified panel of 20 scenarios per atlas (19 leave-one-site-out subsets plus one low-motion subset), under the clustered test. At H_1, STRAND detects the effect with $p_{\mathrm{cl}} < 0.05$ and HR confidence interval excluding 1 in 60/60 scenarios, with tight within-atlas HR ranges (Schaefer-400 1.130–1.161, CC200 1.087–1.121, AAL90 1.082–1.110). Holding out NYU, the largest site, leaves the AAL90 H_1 result intact ($p_{\mathrm{cl}} = 0.0025$, HR 1.089 [1.031, 1.149]). At H_0, Schaefer-400 detects the effect in 20/20 scenarios (HR 0.743–0.866), while CC200 and AAL90 give 0/20, consistent with the primary analysis.

Effect-size attenuation across parcellation resolution. The HR magnitude at H_1 shrinks monotonically with parcellation coarseness: 1.13–1.16 (Schaefer-400, 400 ROIs), 1.09–1.12 (CC200, 200 ROIs), 1.08–1.11 (AAL90, 90 ROIs), with the Schaefer-400 window above the two coarser ones. The pattern is consistent with finer parcellations resolving more of the topological structure that distinguishes the two groups, attenuating the per-feature hazard ratio at coarser resolutions. This characterisation of the effect-size dependence on parcellation choice is unique to STRAND among the methods compared: the baselines provide p-values that are similarly significant across atlases (all $p < 0.05$ at H_1 on every atlas) but cannot quantify the structural effect-size shrinkage that the survival framework makes explicit.

I Discussion of methodological choices

This appendix discusses several methodological choices that practitioners may wish to revisit, and clarifies the relationship between STRAND and existing curve-based summaries.

I.1 Moments and functionals of persistence.

As a useful consequence of the survival representation, the persistence survival function determines all integrable functionals of persistence through

$$\mathbb{E}[\, g(p)\,] = \int_0^\infty g'(t)\, S(t)\, dt \tag{49}$$

for any absolutely continuous g with $g(0) = 0$. Setting $g(t) = t^k$ gives $\mathbb{E}[p^k] = k \int_0^\infty t^{k-1} S(t)\, dt$; in particular $\mathbb{E}[p] = \int_0^\infty S(t)\, dt$, the mean persistence. Eq. (49) shows that classical scalar summaries of a diagram (mean, variance, total persistence, L^p norms of persistence) are linear functionals of S.

I.2 Why STRAND rather than STRAND-B as default

The left-truncated estimator STRAND-B is the principled choice when the structural object of interest is the conditional death-time distribution given birth, i.e., $S_B(t) = \mathbb{P}(d > t \mid b < t)$. Yet our classification benchmarks (Tables 14–21) consistently show STRAND-B underperforming both STRAND and STRAND-F. This is not a defect of left truncation but a consequence of what left truncation removes from the risk set.

Each diagram contains a long tail of short-lived features (those with persistence below the W_1 noise scale). Under the marginal estimator of Eq. (2), these features enter the risk set at $t = 0$ and contribute to $\hat{S}(t)$ for small t, where their distribution is class-discriminative even when individually they would be matched to the diagonal under W_1. Under STRAND-B, the same features enter the risk set only at $t = b_i$ and exit at $t = d_i$, so they contribute to $\hat{S}_B$ over the narrow interval $[b_i, d_i]$ rather than uniformly across $[0, d_i]$. The marginal estimand thus pools more information per feature than the conditional estimand, even though the latter is closer to the underlying generative process. This trade-off favours the marginal in finite samples, which is the regime relevant for downstream classification.

For applications where the conditional distribution is itself the scientific object of interest — e.g., comparing how the rate of feature death changes with filtration scale across cohorts — STRAND-B remains the appropriate variant. We retain it as a documented option rather than the default.

I.3 What the persistence test and the birth test detect

Table 23 compares four tests over 500 replications with a diagram-level permutation null, on the six manifold configurations of Section 4.1: the log-rank test on persistence (STRAND), the same test on the birth survival function, and STRAND-F combining the two by Bonferroni correction or by a single statistic with one shared permutation null. The persistence test is blind to births by construction, since $\hat{S}$ is a functional of the lifetime multiset alone, and the birth test is blind to lifetimes; STRAND-F detects both and stays calibrated under the null.

Table 23: Rejection rates at $\alpha = 0.05$, 500 replications, diagram-level permutation.

Scenario	Persistence	Birth	STRAND-F (Bonf.)	STRAND-F (combined)
Identical diagrams	0.056	0.046	0.052	0.044
Same lifetimes, different births	0.044	1.000	1.000	1.000
Different lifetimes, same births	1.000	0.050	1.000	1.000

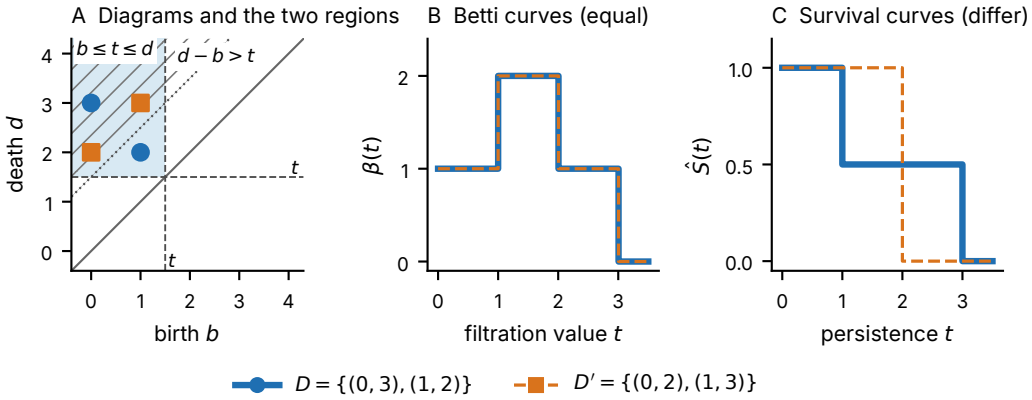


Figure 7: Two diagrams with identical Betti curves and different persistence survival functions. A: $D = \{(0,3),(1,2)\}$ and $D' = \{(0,2),(1,3)\}$, with the quadrant $\{b \le t \le d\}$ counted by $\beta(t)$ (shaded) and the half-plane $\{d-b > t\}$ counted by $\hat{S}(t)$ (hatched), here at $t = 1.5$. B: the Betti curves coincide. C: the survival curves differ, since the lifetimes are $\{3,1\}$ and $\{2,2\}$; $W_1(D,D') = 2$.

I.4 Geometric distinction from Betti curves

Both STRAND and the Betti curve are real-valued summaries of a persistence diagram derived from indicator counts, but they slice the diagram along different axes. The Betti curve at filtration value t counts features alive at scale t:

$$\beta(t) = \#\{i : b_i \le t \le d_i\}, \tag{50}$$

which corresponds geometrically to the quadrant $\{b \le t \le d\}$, whose corner slides along the diagonal as t increases. The persistence survival function counts features whose lifetime exceeds a threshold:

$$\hat{S}(t) = \frac{1}{n}\#\{i : d_i - b_i > t\}, \tag{51}$$

which corresponds geometrically to the half-plane above the line $d - b = t$, i.e., a diagonal sweep moving away from the diagonal as t increases. The two summaries operate on different axes of the diagram (filtration parameter for β, persistence value for $\hat{S}$), and consequently respond differently to translations $(b_i, d_i) \mapsto (b_i + c, d_i + c)$: $\beta(t)$ shifts along its argument, whereas $\hat{S}$ is invariant.

This invariance is a choice of estimand: STRAND compares lifetime distributions, and STRAND-B applies the same survival machinery to the pairs (b_i, d_i) through left truncation (Section 3.3).

Filtration curves. O'Bray et al. [2021] generalise the Betti curve on graphs: a graph descriptor, such as the number of connected components or a node-label histogram, is evaluated on each subgraph of an edge-weight filtration, which gives a step function of the threshold. The component count recovers β_0. Filtration curves are therefore indexed by the filtration parameter, like β, and the comparison above carries over. They also bypass the diagram. This lets them record descriptors that persistent homology does not, but counts along the filtration do not determine lifetimes (Figure 7), so the log-rank test and hazard ratio of STRAND have no direct counterpart for them.

On hypothesis testing through Betti curves. A discretised Betti curve can be plugged into a generic permutation test [Islambekov and Pathirana, 2024, Robinson and Turner, 2017], as can

any vectorisation including STRAND's. The contribution of STRAND is not that hypothesis testing is unavailable for Betti curves; it is that the survival representation natively yields a hypothesis test (the log-rank test) along with an interpretable effect size (the hazard ratio) and per-scale localisation (the Kaplan-Meier curve), without requiring an external permutation wrapper.

I.5 When proportional hazards fails

The hazard-ratio interpretation of the log-rank test relies on the proportional hazards (PH) assumption: $\lambda_B(t) = \mathrm{HR} \cdot \lambda_A(t)$ for all t. The validity of the log-rank test as a Type I-controlled procedure does not depend on PH; only the interpretation of $\widehat{\mathrm{HR}}$ as a single summary effect does. Three regimes are worth distinguishing:

1. **PH holds.** Schoenfeld residuals [Schoenfeld, 1982] fail to reject PH; $\widehat{\mathrm{HR}}$ is a meaningful summary; we report it with bootstrap CI.
2. **PH violated, monotone deviation.** Hazard ratios cross or change monotonically over the persistence axis. The log-rank test remains valid; $\widehat{\mathrm{HR}}$ becomes an average of time-varying ratios. We recommend reporting the median persistence difference $\Delta_{\mathrm{med}} = m_A - m_B$ (Section 3.2) as the primary summary, with the Kaplan-Meier curves shown for transparency.
3. **PH violated, non-monotone deviation.** Time-varying coefficient models [Handorf et al., 2024] or stratified log-rank tests can be applied. These are standard tools in survival analysis and apply directly to the persistence survival framework without modification.

The hazard ratio is a single summary only under proportional hazards. A feature-level Schoenfeld test, which ignores within-diagram dependence, does not reject the assumption for 24 of 62 comparisons across our experiments (3 of 12 synthetic signal-versus-noise contrasts, where the alternative is a mixture and non-proportionality is expected; 18 of 44 graph-benchmark contrasts; 3 of 6 ABIDE contrasts). That test is anti-conservative under the design effects of Section 4.3: with the subject as the unit (cluster-robust Schoenfeld test and a subject-bootstrap of its slope, Appendix H), proportional hazards is retained for all three ABIDE H_1 contrasts and for Schaefer-400 at H_0 ($p = 0.57$, against 1.1×10^{-6} at the feature level), and rejected only for AAL90 at H_0 ($p = 0.027$) and marginally for CC200 at H_0 ($p = 0.066$, slope interval excluding 0). For those two cells, both null results, the median persistence difference is 0.006 $[-0.013, 0.027]$ and 0.006 $[-0.013, 0.025]$, with subject-bootstrap intervals covering 0. Table 5 reports the hazard ratio in every cell, and for the two cells where the subject-level test rejects the assumption the median difference above is the summary to read.

I.6 On test hypersensitivity from large feature pools

Pooling persistence values across many diagrams can produce n_{eff} in the thousands, raising concern that the log-rank test will detect statistically significant but practically negligible differences. This is a general property of well-powered tests [Wasserstein and Lazar, 2016], not specific to STRAND.

The persistence survival framework is, however, equipped to diagnose this regime explicitly. A significant p-value with $\widehat{\mathrm{HR}} \approx 1$ (CI tight around 1) indicates a detectable but negligible effect, which is exactly the diagnostic that scalar-p-value-only methods (e.g., the permutation test of Robinson and Turner [2017]) cannot provide. Practitioners can read off the magnitude alongside the significance, and report findings using both. Our power analysis (Appendix F, Figure 6) corroborates that the test is not spuriously sensitive: at $\mathrm{HR} = 1$ the empirical rejection rate is at the nominal α level (Table 10), and the V-shaped power curve confirms that rejections concentrate where the true HR deviates from 1.

I.7 Limitations

STRAND summarises a diagram by its lifetime distribution, so diagrams related by a diagonal translation are indistinguishable to it, and the birth and lifetime marginals used by STRAND-F do not determine their joint distribution (Section 3.3). Against tests that compare full diagrams, the lifetime marginal costs power at weak contrasts (Appendix F). The hazard ratio is a single summary only under proportional hazards; when that fails we report median persistence differences and Kaplan–Meier curves (Appendix I.5). Inference is asymptotic in the number of diagrams; with few diagrams

per group we use the diagram-level permutation test, which cannot reject at $\alpha = 0.05$ below four diagrams per group. The stability guarantees hold for a fixed grid and are integrated (Section 3.1). On graphs we use the degree filtration for comparability with prior work, which limits what any of the compared summaries can express on attributed graphs.

What the lifetime marginal discards. $\hat{S}$ depends on a diagram only through its persistence values, so it is blind to diagonal translations and to any rearrangement of births that keeps the lifetimes, whereas W_1 is not. No fixed-dimensional summary reflects W_1 from below: a bi-Lipschitz embedding of persistence diagrams into a finite-dimensional Hilbert space does not exist, even for diagrams of bounded cardinality [Carrière and Bauer, 2019], and the Betti-curve pair of Section 3.1 (Figure 7, Appendix I.4) has identical curves at $W_1 = 2$. On our graph benchmarks, the cost of the invariance can be measured with `STRAND`-F, which adds the birth marginal: the TU Dortmund average moves from 60.7% to 60.9% with SVM and from 60.5% to 61.8% with Random Forest, with no per-dataset change above one standard deviation (Appendix G). Since the two marginals do not determine the joint, this is a lower bound on what the invariance costs.

J Corrections relative to the submitted version

Four points were corrected after review. For each, the superseded result is shown next to the corrected one.

Variance of the two-sample test. The submitted version computed the log-rank test with the hypergeometric variance $\sum_j V_j$ of Eq. (6), which treats pooled persistence values as independent. Features of one diagram are dependent, and in the multi-diagram design of Table 1 that variance rejects a true null up to 67% of the time; the clustered variance of Eq. (7) is calibrated in the same design. Both columns are printed in Table 1. The submitted calibration compared one diagram per group, where the feature is the sampling unit and the classical variance is the appropriate one; it remains valid for that design and is kept in Appendix D. The hazard ratio is now weighted by inverse diagram size so that it targets the estimand of Eq. (3); the unweighted values differ by at most 0.003 (Appendix H).

Power from two diagrams per group. The submitted version reported power 0.91 at $n = 2$ diagrams per group. With two diagrams per group an exact permutation test has six label assignments and cannot reject at $\alpha = 0.05$; the reported value reflected the number of features, not evidence about the two groups. Table 24 shows the superseded values next to the clustered and exact tests. Power is now reported from $n = 5$, where the clustered and the exact permutation test agree, and the exact test is recommended below that.

Table 24: Power at $\alpha = 0.05$, dimension 1, 4D torus, as a function of diagrams per group. Naive: submitted test; Clustered: Eq. (7); Exact: diagram-level permutation. Floor: smallest attainable two-sided permutation p-value, $2/\binom{2n}{n}$.

n per group	Floor	Naive (submitted)	Clustered	Exact
2	0.333	0.91	0.23	0.00
3	0.100	0.98	0.90	0.00
5	0.008	1.00	1.00	1.00
10	< 0.001	1.00	1.00	1.00

ABIDE. The submitted Table 5 used a site-stratified classical log-rank test. Table 25 shows that column, the unstratified classical test and the corrected tests. At CC200 in dimension 0 the unstratified classical test rejects ($p = 0.039$) where the clustered test and the exact permutation retain the null (0.743); every dimension-1 result and the Schaefer-400 dimension-0 result stand under the corrected tests, with subject-level confidence intervals.

Stability. The submitted version stated, after Theorem 3.1, that both terms of the bound vanish as $W_1(D, D') \to 0$, and derived from it a corollary that $\|\phi_S(D_n; \mathbf{t}) - \phi_S(D; \mathbf{t})\|_\infty \to 0$ whenever $W_1(D_n, D) \to 0$ at equal cardinality. Both are false for evaluation on a fixed grid: the gap δ is itself

Table 25: ABIDE p-values under the submitted and the corrected tests.

Dim	Atlas	Site-stratified classical (submitted)	Unstratified classical	Clustered	Exact permutation
H_0	Schaefer-400	0.006	< 0.001	3.1×10^{-4}	< 0.001
	CC200	0.801	0.039	0.743	0.743
	AAL90	0.622	0.800	0.504	0.498
H_1	Schaefer-400	< 0.001	< 0.001	2.6×10^{-12}	< 0.001
	CC200	0.001	< 0.001	1.6×10^{-6}	< 0.001
	AAL90	0.021	< 0.001	3.4×10^{-4}	0.002

at most $2\,W_1(D, D')$, so the second term is at least $1/\min(n, m)$ as soon as two matched persistence values differ, and a single persistence value crossing a grid point moves one coordinate by $1/n$ at arbitrarily small W_1. The inequality of Theorem 3.1 is unchanged. The vanishing claim and the corollary are replaced by the integrated bound of Proposition 3.2 and the eventual-equality statement of Proposition 3.3, and the vectorisation is no longer described as pointwise Wasserstein-stable.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state three contributions: (i) a two-sample test on the persistence-lifetime distributions of collections of diagrams, with the diagram as the sampling unit (Section 3.2, Section 4.1), (ii) hazard ratios with subject-level confidence intervals, median persistence differences and Kaplan–Meier curves (Sections 3.2–4.3), and (iii) a vectorisation with an integrated stability bound (Section 3.3, Section 4.2). Each claim is supported by the section it points to. Theorem 3.1 and Propositions 3.2 and 3.3 are proven in Appendix C, and Proposition 3.4 restates the power formula of Schoenfeld [1981].

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: A Limitations paragraph in Appendix I.6 cover the invariance of the lifetime marginal to diagonal translations and its cost in power against full-diagram tests (Appendix F.1), the asymptotic nature of the clustered test with the exact permutation test recommended below five diagrams per group, the proportional-hazards condition for the hazard ratio (Appendix I.4), the integrated rather than pointwise stability guarantee, and the degree filtration on graphs. Corrections relative to the submitted version are listed in Appendix J.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.

- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Theorem 3.1 and Propositions 3.2 and 3.3 state their assumptions on the diagrams, the grid and the optimal transport; complete proofs are in Appendix C.1 with supporting Lemmas C.6 to C.8. Proposition 3.4 states its assumptions (independent pooled persistence values, proportional hazards, equal allocation) and cites the closed-form result; Section 3.2 and Appendix E address within-diagram dependence.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Appendix D gives the simulation design (manifolds, noise models, mixture null, replications, seeds); Appendix E the calibration protocol; Appendix G.2 the cross-validation protocol and hyperparameter grids for every method; Section 4.3 the ABIDE preprocessing (correlation-to-distance transform, Vietoris–Rips truncation, quality-control rule). All numbers are reproduced by the released code.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Code at https://github.com/BorgwardtLab/strand-tda, with a README, installation instructions and scripts for every table and figure. Public datasets are accessed through their original repositories (TU Dortmund via torch_geometric, the flooder benchmarks via the flooder package, ABIDE I via the Preprocessed Connectomes Project).

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 4.2 and Appendix G.2 specify the classifiers (SVM for every method, Random Forest for every vectorisation), grid sizes, splits (stratified 10-fold for graphs, the 10 predefined splits for point clouds), and hyperparameter grids with selection on the training portion of each split. Section 4.3 specifies the inference protocol (clustered log-rank with

the subject as cluster, diagram-level permutation with B = 1,000, 1,000 subject-bootstrap resamples) and the three comparators with their permutation counts. Table 4 states its timing protocol.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Type I error and power are reported with Wilson 95% intervals over 1,000 or 200 replications (Tables 1, 2, 7, 8, 10). Classification results report mean and standard deviation over folds (Appendix G) and pairwise Wilcoxon signed-rank tests (Section 4.2). Hazard ratios carry 95% intervals from 1,000 bootstrap resamples of subjects, and ABIDE p-values are reported from both the clustered test and an exact subject-level permutation (Table 5).

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: All experiments ran on CPU nodes (AMD EPYC 9654) of the Max Planck Computing and Data Facility; Table 4 reports per-method runtimes under a stated protocol, and no experiment requires a GPU.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.

- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have read the NeurIPS Code of Ethics. The work uses publicly available datasets (TU Dortmund, flooder benchmarks, ABIDE I) under their respective release terms, contains no human-subjects experimentation, and does not involve any release of high-risk models. ABIDE I data is de-identified and was originally collected under IRB protocols at the contributing sites; our use is purely secondary analysis of the public release.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: `STRAND` is a methodological contribution to topological data analysis. Positive impacts include enabling rigorous, effect-size-aware statistical inference in domains where TDA is increasingly used (drug discovery, biomedical imaging, materials science, neuroimaging), supporting better-grounded scientific claims and meta-analysis. Potential negative impacts are limited: as a generic statistical method, `STRAND` does not enable applications that were not already possible with existing PD inference methods, but as with any inferential tool it could be misapplied (e.g., reporting effect sizes outside the proportional hazards regime, or treating exploratory findings as confirmatory). The paper recommends proportional hazards diagnostics (Schoenfeld residuals) and reports non-parametric alternatives when assumptions fail.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.

- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release pretrained models, generative systems, or scraped data. `STRAND` is an algorithmic contribution implemented on top of standard libraries. No high-risk-of-misuse asset is being released.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All datasets and software libraries used are cited at first use. TU Dortmund graph datasets [Morris et al., 2020] are released for academic use. The flooder package and benchmarks [**?**] are released under their NeurIPS 2025 release terms. ABIDE I [Heinsfeld et al., 2018] is released under the ABIDE Preprocessed Initiative terms (Creative Commons Attribution-NonCommercial-Share Alike 3.0 Unported License). Software libraries (`GUDHI, Ripser, lifelines, scikit-learn`) are used under their respective open-source licenses.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The STRAND implementation is released at the URL above, with a README documenting installation, dependencies, supported input formats (GUDHI, Ripser, Giotto-TDA, dict) and usage examples for the vectoriser and the two-sample test.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A] s

Justification: The paper does not involve crowdsourcing or new research with human subjects. ABIDE I is a publicly released, de-identified dataset analysed under its public-release terms.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: No new human-subjects research was conducted. ABIDE I data was collected under IRB protocols at the original contributing sites and released publicly in de-identified form; secondary analysis of this released data does not require additional IRB review.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs are not a component of the core methodology. Any use of LLMs in the preparation of this manuscript was limited to writing and editing assistance.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.