Bayesian Analysis (2021) **TBA**, Number TBA, pp. 1–26

Bayesian Topological Learning for Classifying the Structure of Biological Networks

Vasileios Maroulas*, Cassie Putman Micucci†, and Farzana Nasrin‡

Abstract. Actin cytoskeleton networks generate local topological signatures due to the natural variations in the number, size, and shape of holes of the networks. Persistent homology is a method that explores these topological properties of data and summarizes them as persistence diagrams. In this work, we analyze and classify simulated actin filament networks by transforming them into persistence diagrams whose variability is quantified via a Bayesian framework on the space of persistence diagrams. The proposed generalized Bayesian framework adopts an independent and identically distributed cluster point process characterization of persistence diagrams and relies on a substitution likelihood argument. This framework provides the flexibility to estimate the posterior cardinality distribution of points in a persistence diagram and their posterior spatial distribution simultaneously. We present a closed form of the posteriors under the assumption of a Gaussian mixture and binomial for prior intensity and cardinality respectively. Using this posterior calculation, finally, we implement a Bayes factor algorithm to classify simulated actin filament networks and benchmark it against several state-of-the-art classification methods.

Keywords: Bayesian inference and classification, intensity, cardinality, marked point processes, topological data analysis.

MSC2020 subject classifications: Primary 62F15, 60G55, 62-07; secondary 62P10.

1 Introduction

The actively functioning transportation of various particles through intracellular movements is a vital process for cells of living organisms (Porter and Day (2016)). Such transportation must be intricately organized due to the tightly packed nature of the interior of a cell at the molecular level (Breuer et al. (2017)). The actin cytoskeleton, which consists of actin filaments cross-linking with myosin motor proteins along with other pertinent binding proteins, is an important component in plant cells that determines the structure of the cell and provides transport of cellular components (Freedman et al. (2017); Breuer et al. (2017)). Although researchers have investigated the molecular features of actin cytoskeletons (e.g., Staiger et al. (2000); Shimmen and Yokota (2004); Freedman et al. (2017); Mlynarczyk and Abel (2019)), the underlying process that determines their structures and how these structures are linked to intracellular transport remains undetermined (Thomas et al. (2009); Madison and Nebenführ (2013)). A crucial step to understand this transport is to define quantitative measures of the actin

*Department of Mathematics, University of Tennessee, Knoxville, TN
†Eastman Chemical Company, Kingsport, TN
‡Department of Mathematics, University of Hawai'i at Mānoa, Honolulu, HI, fnasrin@hawaii.edu

 https://doi.org/10.1214/21-BA1270

cytoskeleton's structure, and understand the different structural networks of filaments on which the organelles are moving. However there is not a method for either fully depicting the characteristics of networks.

Researchers are interested in developing and using quantitative tools to capture actin filament structures. The analysis of skewness in the pixel intensities of microscopic images is performed to quantify the actin cytoskeleton bundles and density in Higaki et al. (2010). Polymer network-based models are studied and classified to identify how actin cytoskeleton structure can be efficiently represented in Banerjee and Park (2015). On the other hand, from a closer look, the inherent variation in size, density, and positioning of actin filaments yields topological signatures in the cytoskeleton's network, (Tang et al. (2014)). In this article, we develop a fully data-driven Bayesian topological learning method, which could aid researchers by providing a pathway to predict cytoskeleton structural properties by classifying simulated actin filament networks to identify the effect of the number of cross-linking proteins on the network. With more cross-linking proteins available, the cell has networks with many binding locations, which create larger loops within the whole structure of the actin cytoskeleton. Topological data analysis (TDA) can be viewed as a dimensionality reduction method that allows us to map data from high dimensional space to a lower dimensional space. When viewed through the lens of topology, these networks show dissimilarity due to the presence and size of loops. Differentiating between the empty space and the connectedness of these networks allows us to create an accurate classification rule using topological methods. Although we focus on our analysis to the classification of actin filament networks, the topological Bayesian framework could be generalized to other data sets.

Persistent homology is a powerful TDA tool that provides a robust way to model the topology of data and summarizes salient features with persistence diagrams (PDs). These diagrams are multisets of points in the plane, each point representing a homological feature whose time of appearance and disappearance is contained in the coordinates of that point (Edelsbrunner and Harer (2010)). Persistent homology has proven to be promising in a variety of applications such as shape analysis Patrangenaru et al. (2018), image analysis Guo et al. (2018), neuroscience Sizemore et al. (2018); Biscio and Møller (2019); Nasrin et al. (2019), sensor networks Dłotko et al. (2012); Carlsson and de Silva (2010), biology Sgouralis et al. (2017); Maroulas and Nebenführ (2015); Mike et al. (2016); Nicolau et al. (2011), dynamical systems Khasawneh and Munch (2016), action recognition Venkataraman et al. (2016), signal analysis Marchese and Maroulas (2018, 2016), chemistry and material science, Kimura et al. (2018); Maroulas et al. (2020); Townsend et al. (2020), and genetics Humphreys et al. (2019).

While there are several methods present in the literature to compute PDs, we choose geometric complexes that are typically used for applications of persistent homology to data analysis; see Edelsbrunner and Harer (2010) and references therein. The homological features in PDs have no intrinsic order, implying that they are sets as opposed to vectors. Due to this, the utilization of PDs in machine learning algorithms is not straightforward. Some researchers map the PDs into Hilbert spaces to adopt traditional machine learning tools (see e.g., Di Fabio and Ferri (2015); Turner et al. (2014); Adams et al. (2017); Bubenik (2015); Reininghaus et al. (2015)). Direct use of PDs for statistical inference and classification has been developed by several authors such as

Maroulas et al. (2019, 2020); Marchese and Maroulas (2018); Bobrowski et al. (2017); Fasy et al. (2014); Mileyko et al. (2011); Robinson and Turner (2017).

In this paper, we quantify the variability of PDs through a novel Bayesian framework by considering PDs as a collection of points distributed on a pertinent domain space, where the distribution of the number of points is also an important feature. This setting leads us to view a PD through the lens of an independent and identically distributed (i.i.d.) cluster point process (PP) (Daley and Vere-Jones (1988)). An i.i.d. cluster PP consists of points that are i.i.d. according to a probability density but have an arbitrary cardinality distribution. For example, an i.i.d. cluster PP is reduced to the classical Poisson PP if the points in a PD are spatially distributed according to a Poisson distribution. The study in Maroulas et al. (2020) implicitly estimates the cardinality of a PD by integrating the intensity of a Poisson PP. The framework of Maroulas et al. (2020) also yields that the variance is equal to the mean and leads to an estimation of cardinality with high variance whenever the number of points in a PD is high. However, modeling PDs as i.i.d. cluster PPs allows us to estimate the intensity and the cardinality component of the distribution simultaneously. This is very critical as the importance of cardinality in PDs has been underlined in problems related to statistics and machine learning Fasy et al. (2014); Kerber et al. (2017).

Our Bayesian framework quantifies prior uncertainty with given intensity and cardinality for an i.i.d. cluster PP. The likelihood in our model represents the level of belief that observed diagrams are representative of the entire population and are defined through marked point processes (MPPs). A central idea of this paper is to develop posterior distributions of the spatial configuration of points on persistence diagrams and their associated number instead of the point clouds in the data generating space. The persistence diagrams summarize their topology which in turn is employed in the classification algorithm. By viewing point clouds through their topological descriptors, the proposed framework can reveal essential shape peculiarities latent in the point clouds. Our Bayesian method adopts a substitution likelihood technique by Jeffreys in Jeffreys (1961) instead of considering the full likelihood for the point cloud. Due to the nature of PDs, an observed PD contains points that correspond to the latent topology in the underlying data as well as points that solely arise due to noise in the data. Our Bayesian model addresses instances of noise by means of an i.i.d. cluster PP. In particular, we are able to quantify the uncertainty with an estimated intensity and cardinality using the i.i.d. cluster PP. This framework estimates the posterior cardinality and intensity simultaneously, which provides a complete knowledge of the posterior distribution.

Another key contribution of this paper is the derivation of a closed form of the posterior intensity, which relies on Gaussian mixture densities for prior intensities *and* a closed form for the posterior cardinality, which uses binomial priors. The direct benefits of this closed form solution of the posterior distribution are two-fold: (i) it demonstrates the computational tractability of the proposed Bayesian model and (ii) it provides a means to develop a robust classification scheme through Bayes' factors. Another computational benefit of these closed forms is the quantification of the intensity of the unexpected PP by means of an exponential density. The exponential density is an ideal choice because (i) it gives a natural intuition of the unexpected (noise) features, and

(ii) it provides a more computationally automatic approach as we only need to modify one parameter. This Bayesian paradigm provides a method for the classification of actin filament networks in plant cells that captures their distinguishing topological features.

Overall, the contributions of this work are:

1. A generalized Bayesian framework that simultaneously estimates the spatial and the cardinality distribution of PDs using i.i.d. cluster PPs.
2. A general closed form expression of both the posterior spatial distribution and the posterior cardinality distribution of PDs.
3. A Bayesian classification algorithm for actin filament networks of plant cells that directly incorporates the variations in topological structures of those networks such as number and size of loops.

This paper is organized as follows. Section 2 provides a brief overview of PDs and PPs. In Section 3, we establish the Bayesian framework for PDs and provide the update formulas for intensity and cardinality. Then Subsection 3.1 introduces a closed form representation of the posterior intensity and cardinality utilizing Gaussian mixture models and binomial distributions respectively. Detailed demonstrations of this closed form estimation are presented in Subsection 3.2. To assess the capability of our Bayesian method, we investigate a problem of classifying filament networks of plant cells in Section 4. Finally, we end with a discussion in Section 5. We delegate all of the proofs, as well as some definitions, lemmas, and notations required for the proofs to the supplementary materials (Maroulas et al., 2021).

2 Preliminaries

We begin by discussing the necessary background for generating Bayesian models for PDs. In Subsection 2.1, we briefly review simplicial complexes, the building blocks for constructing PDs. Pertinent definitions, theorems, and some basic facts about i.i.d. cluster point processes (PPs) are discussed in Subsection 2.2.

2.1 Persistence Diagrams

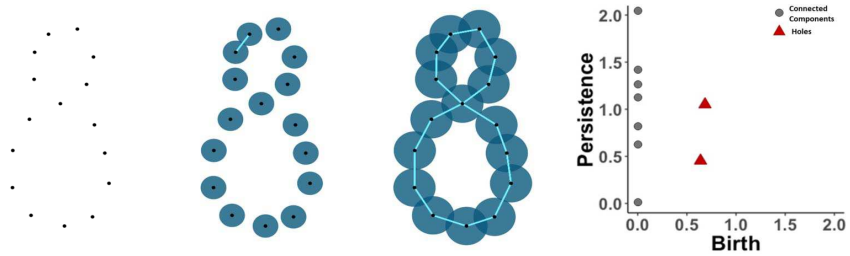
Definition 2.1. The convex hull of a finite set of points $\{x_i\}_{i=1}^n$ is given by $\sum_{i=1}^n \alpha_i x_i$, where $\alpha_i \geq 0$ for all i and $\sum_{i=1}^n \alpha_i = 1$.

Definition 2.2. The set of points $\{x_i\}_{i=1}^n$ is affinely independent if whenever $\sum_{i=1}^n \alpha_i x_i = 0$ and $\sum_{i=1}^n \alpha_i = 0$, then $\alpha_i = 0$ for all i.

Definition 2.3. A k-simplex, $[x_0, \ldots, x_k]$ is a collection of $k+1$ affinely independent elements along with their convex hull. The faces of a k-simplex are the $(k-1)$-simplices spanned by subsets of $\{x_0, \ldots, x_k\}$.

Definition 2.4. A simplicial complex σ is a collection of simplices such that for every set A in σ and every nonempty set $B \subset A$, we have that B is in σ.

Definition 2.5. The Vietoris-Rips complex for threshold $\epsilon > 0$, denoted VR(ϵ), is the abstract simplicial complex determined in the following way: a k-simplex with vertices given by $k+1$ points in X is included in $VR(\epsilon)$ whenever $\epsilon/2$ balls placed at the points all have pairwise intersections.



(a) (b) (c) (d)

Figure 1: (a) An underlying dataset of points. (b) The Vietoris-Rips complex consisting of the points and the light blue line segment. (c) The Vietoris-Rips complex consisting of the points and line segments that now form an "eight" shape, which has two 1-dimensional holes. (d) The tilted PD for connected components and holes.

Formally, for each homological dimension, a PD is a multiset of points (b, d), where b is the radius in the Vietoris-Rips complex at which a homological feature is born and d is the radius at which it dies. Intuitively, the homological features represented in a PD are connected components or holes of different dimensions. To illustrate the Vietoris-Rips complexes we present a toy example in Figure 1 by considering an "eight" shape in (a). The algorithm starts by taking into account circles with increasing radii (at each algorithmic step) centered at each data point. As the "resolution" of the data changes by increasing the radii, homological features emerge or disappear by examining if two or more circles intersect. For example, Figure 1(b) presents a scenario where only two circles intersect. While the circles centered at each data point grow and more connected components are generated, holes and voids may be also created (see Figure 1 (c)). Eventually, they get filled due to increasing the radii, and the process ends when all circles intersect. The algorithmic results are summarized in a persistence diagram. Each point in a persistence diagram (shown as red triangles in Figure 1 (d)) is referred as (b, p) in $\mathbb{W} := \{(b, p) \in \mathbb{R}^2 |\ b, p \geq 0\}$ where b is the birth scale and $p = d - b$ is the persistence value.

2.2 I.I.D. Cluster Point Processes

This section contains basic definitions and fundamental theorems related to i.i.d. cluster PPs. Detailed treatments of i.i.d. cluster PPs can be found in Daley and Vere-Jones (1988) and references therein. Throughout this section, we let $\mathbb{X}$ be a Polish space and $\mathcal{X}$ be its Borel σ-algebra.

Definition 2.6. A finite point process $(\{\rho_n\}, \{\mathbb{P}_n(\bullet)\})$ consists of a cardinality distribution ρ_n with $\sum_{n=0}^{\infty} \rho_n = 1$ and a symmetric probability measure $\mathbb{P}_n$ on $\mathcal{X}^n$, where $\mathcal{X}^0$ is the trivial σ-algebra.

To sample from a PP, first one draws an integer n from the cardinality distribution ρ_n. Then the n points $(x_1, \ldots, x_n)$ are spatially distributed according to a draw from $\mathbb{P}_n$. Since PPs model unordered collections of points, we need to ensure that $\mathbb{P}_n$ assigns equal weights to all $n!$ permutations of $(x_1, \ldots, x_n)$. The requirement in Definition 2.6 that $\mathbb{P}_n$ is symmetric guarantees this. A natural way to work with random collections of points is the Janossy measure, which combines the cardinality and spatial distributions, while disregarding the order of the points.

Definition 2.7. For disjoint rectangles $A_1, \ldots, A_n$, the Janossy measure for a finite point process is given by $\mathbb{J}_n(A_1 \times \cdots \times A_n) = n! \rho_n \mathbb{P}_n(A_1 \times \cdots \times A_n)$.

Definition 2.8. An i.i.d. cluster PP Ψ is a finite PP on the space $(\mathbb{X}, \mathcal{X})$ which has points that: (i) are located in $\mathbb{X} = \mathbb{R}^d$, (ii) have a cardinality distribution ρ_n with $\sum_{n=0}^{\infty} \rho_n = 1$, and (iii) are distributed according to some common probability measure $F(\cdot)$ on the Borel set $\mathcal{X}$.

We consider Janossy measures for the point process Ψ, $\mathbb{J}_n^{\Psi}$, that admit densities j_n with respect to a reference measure on $\mathbb{X}$ due to their intuitive interpretation. In particular, for an i.i.d. cluster PP Ψ, if $F(A) = \int_A f(x)dx$ for any $A \in \mathcal{X}^n$, then $j_n(x_1, \ldots, x_n) = \rho_n n! f(x_1) \cdots f(x_n)$ determines the probability density of finding the n points at their respective locations according to F. The $n!$ term gives the number of ways the points could be at these positions. For a finite intensity measure Λ on $\mathbb{X}$ that admits the density λ, we also have $f(x) = \frac{\lambda(x)}{\Lambda(\mathbb{X})}$. The intensity is the point process analog of the first order moment of a random variable. Precisely, the intensity density $\lambda(x)$ is the density of the expected number of points per unit volume at x. Hereafter, we sufficiently characterize our i.i.d. cluster PPs with intensity and cardinality measures. Next, we define the marked PP, which provides a formulation for the likelihood model used in our Bayesian setting. Let $\mathcal{M}$ be a Polish space that represents the mark space, and let its Borel σ-algebra be $\mathcal{M}$.

Definition 2.9. Suppose $\ell : \mathbb{X} \times \mathbb{M} \to \mathbb{R}^+ \cup \{0\}$ is a function satisfying: 1) for all $x \in \mathbb{X}$, $\ell(x, \bullet)$ is a probability measure on $\mathbb{M}$, and 2) for all $B \in \mathcal{M}$, $\ell(\bullet, B)$ is a measurable function on $\mathbb{X}$. Then, ℓ is a stochastic kernel from $\mathbb{X}$ to $\mathbb{M}$.

Definition 2.10. A marked i.i.d. cluster PP (Ψ, Ψ_M) is a finite PP on $\mathbb{X} \times \mathbb{M}$ such that: (i) $\Psi = (\{\rho_n\}, \{\mathbb{P}_n(\bullet)\})$ is an i.i.d. cluster PP on $\mathbb{X}$, and (ii) for a realization $(\mathbf{x}, \mathbf{m}) \in \mathbb{X} \times \mathbb{M}$, the marks m_i of each $x_i \in \mathbf{x}$ are drawn independently from a given stochastic kernel $\ell(\bullet|x_i)$.

Remark 1. A marked point process (Ψ, Ψ_M) is a bivariate PP where one point process is parameterized by the other. Therefore, if the cardinalities of $\mathbf{x}$ and $\mathbf{m}$ are equal, then the conditional density for $\mathbf{m}$ is $\ell(\mathbf{m}|\mathbf{x}) = \frac{1}{n!} \sum_{\pi \in \mathcal{S}_n} \prod_{i=1}^{n} \ell(m_i|x_{\pi(i)})$, where $\mathcal{S}_n$ is the set of all permutations of $(1, \ldots, n)$. Otherwise, the density can be taken as 0.

The final two definitions we need to construct the Bayesian theorem are the probability generating functional (PGFL) and elementary symmetric function. The PGFL is a point process analogue of the probability generating function (PGF) of random variables. Intuitively, the point process can be characterized by the functional derivatives of the PGFL (Moyal (1962)). The other necessary definitions and theorems related to the PGFL, which will be heavily used in the proof of Theorem 3.1 are provided in Section 1 of the supplementary materials.

Definition 2.11. Let Ψ be a finite PP on $\mathbb{X}$ and $\mathcal{H}$ be the Banach space of all bounded measurable complex valued functions ζ on $\mathbb{X}$. For a symmetric function, $\zeta(\mathbf{x}) = \zeta(x_1)\cdots\zeta(x_n)$ and $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{X}$, the PGFL of Ψ is given by

$$G(\zeta) = \mathbb{E}\big[\prod_{j=1}^{n} \zeta(x_j)\big] = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\mathbb{X}^n} \left(\prod_{j=1}^{n} \zeta(x_j)\right) \mathbb{J}_n(dx_1 \ldots dx_n). \tag{2.1}$$

The first expression shows the analogy of the PGFL with the PGF, as it is the expectation of the product $\prod_{j=1}^{n} \zeta(x_j)$. Hence, if $\zeta(x_i) = x$, a constant real non-negative number for all x_i, then $G(\zeta)$ takes the form of a PGF $g_N(x) = \sum_{n=0}^{\infty} p_N(n)x^n$, where $p_N(n)$ is the probability distribution of a random $N \in \mathbb{N}_0 = \{0, 1, 2, \cdots\}$.

Remark 2. For an i.i.d. cluster process Ψ the PGFL has the form (Daley and Vere-Jones (1988)):

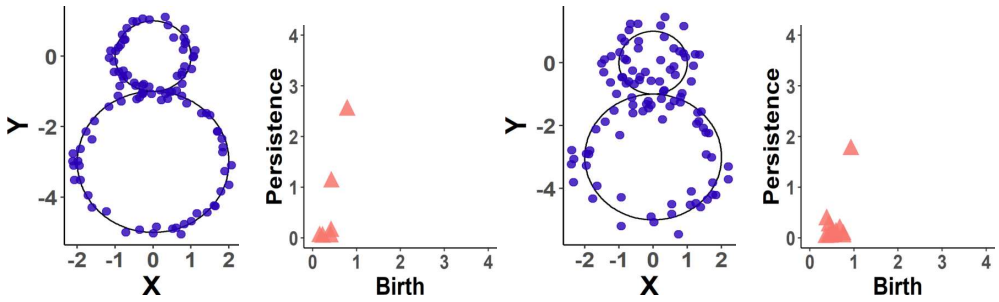
$$G(\zeta) = g_N\Big(\int_{\mathcal{X}} \zeta(x) f(x) dx\Big), \tag{2.2}$$

where g_N is the PGF of the cardinality N, ζ has the same form as in Definition 2.11, and f is the probability density discussed after Definition 2.8.

Definition 2.12. The elementary symmetric function $e_{K,k}$ is given by $e_{K,k}(\nu_1, \cdots, \nu_K) = \sum_{0 \le i_1 < \cdots < i_k \le K} \nu_{i_1} \cdots \nu_{i_k}$ with $e_{0,k} = 1$ by convention.

3 Bayesian Inference

For developing the framework for Bayesian inference, we consider the underlying prior uncertainty of a PD, D_X, generated by an i.i.d. cluster PP, $\mathcal{D}_X$, with intensity $\lambda_{\mathcal{D}_X}$ and cardinality distribution $\rho_{\mathcal{D}_X} = P(|\mathcal{D}_X| = n)$, the probability of the number of elements in the PP $\mathcal{D}_X$ to be equal to n, where $|\cdot|$ denotes the cardinality. To motivate the discussion and develop the intuition about our method, let's consider the point clouds (blue dots in Figure 2 (a) and (c)) are generated from an 'eight' shape (solid line Figure 2 (a) and (c)). A perfect noiseless point cloud yields a 1-dimensional PD with two points whose y-coordinates (persistence values) are much greater than 0. However, noisy point clouds would lead to PDs which may or may not clearly expose the topological 'fingerprint' of prominent points depending on the level of noise. For example, observe in Figure 2 (b) that the two higher persistence points are well separated from the noise points close to the birth axis. But, that's not the case in Figure 2 (d) where only one prominent point is detected To that end, sample persistence diagram from a prior



(a) (b) (c) (d)

Figure 2: (a) and (c) point clouds (blue dots) generated by sampling the 'eight' shape (solid line) and perturbing with Gaussian noise having variances $0.01I_2$ and $0.1I_2$, respectively. (b) and (d) PDs for 1-dimensional features only of (a) and (c), respectively.

which carries the knowledge about topological signatures of the underlying truth would be partially observed. Hence a point x in $\mathcal{D}_X$ would be observed with probability $\alpha(x)$ or vanish with probability $(1-\alpha(x))$ as a result of noise in the data. Consequently, the prior $\mathcal{D}_X$ is decomposed as $\mathcal{D}_X = \mathcal{D}_{X_O} \cup \mathcal{D}_{X_V}$, where $\mathcal{D}_{X_O}$ includes the points that survived and $\mathcal{D}_{X_V}$ otherwise.

Employing the theory of marked PPs, we establish the likelihood model by considering the association with the observed PDs, D_Y, samples of an i.i.d. cluster PP, $\mathcal{D}_Y$, which is decomposed into $\mathcal{D}_{Y_O}$ and $\mathcal{D}_{Y_U}$. $\mathcal{D}_{Y_O}$ consists of the points in the data (persistence diagram) that are linked to $\mathcal{D}_{X_O}$ via a pertinent marked PP with stochastic kernel $\ell(y|x)$ (as in Definition 2.9). Consequently, for the marked PP $(\mathcal{D}_{X_O}, \mathcal{D}_{Y_O})$, the intensity (spatial) likelihood is computed using the stochastic kernel $\ell(y|x)$. The cardinality likelihood is obtained by the conditional distribution of the observed PD D_Y given that there are n points in $\mathcal{D}_X$. Typically, PDs D_Y consist of points that correspond to the latent topology of the point cloud and that generate solely from noise. Consequently, the points that arise from noise fail to associate with the prior and we call them unexpected features. We model such points as generated by an i.i.d. cluster PP $\mathcal{D}_{Y_U}$ with intensity density $\lambda_{\mathcal{D}_{Y_U}}$ and cardinality probability distribution $\rho_{\mathcal{D}_{Y_U}}$.

Figure 3 gives a visual representation to illustrate the contribution of the prior and the observations to the spatial and the cardinality distributions of PDs of the Bayesian framework. For this, we superimpose two PDs: one is a sample from the prior (D_X and shown as triangles) and the other is the observed PD (D_Y and shown as dots) (see Figure 3 (a)). More precisely, the PD of a synthetic point cloud is considered as D_X and the PD of a perturbed version of the point cloud is considered as D_Y. Any arbitrary point x in D_X is equipped with a probability of being observed, which we present using blue (D_{X_O}) and brown (D_{X_V}) otherwise in Figure 3. Presumably, any point $x \in D_{X_O}$ is marked with an observed point in D_{Y_O} via the marked PPs. This implies that for any possible configuration, the number of points in D_{X_O} will be the same as D_{Y_O} as shown in the blue box in Figure 3 (c). Also, we present the unexpected features D_{Y_U}

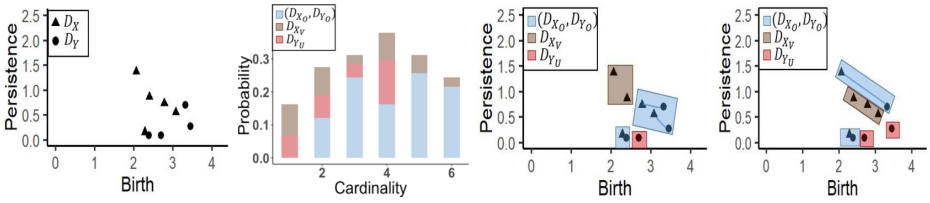


Figure 3: (a) A sample D_X (triangles) from the prior PP $\mathcal{D}_X$ and an observed PD D_Y (dots). (b) is an example of the cardinality distribution of prior and observed PDs which shows some possible configurations of the cardinality probability of $\mathcal{D}_{X_O}, \mathcal{D}_{X_V}, \mathcal{D}_{Y_O}$, and $\mathcal{D}_{Y_U}$. (c-d) are similar color-coded representations of possible configurations of the spatial density for the observed and vanished features in the prior, along with the marked PP $(\mathcal{D}_{X_O}, \mathcal{D}_{Y_O})$ and the unexpected features of the data $\mathcal{D}_{Y_U}$.

(associated to noise) as red in Figure 3. We show different possible scenarios for the relationship of the prior to the data likelihood for the cardinality distribution of PDs in Figure 3 (b). As any point in D_{X_V} has no association with points in the observed PD D_Y, if all of the points in D_X belong to D_{X_V}, in the observed PD we encounter only the unexpected D_{Y_U} (the first bar in Figure 3 (b)). This scenario is possible in the presence of very high noise in data. As some points of D_Y are more likely to be marks than others, we illustrate these instances with different levels for the blue parts of the cardinality bars. The last two bars demonstrate cases where all of the points in D_Y are expected to be marks of the prior features; this is encountered in the presence of very low noise in data.

Figure 3 (c) and (d) show different possible scenarios for the relationship of the prior to the data likelihood for the spatial distribution of points on a PD. The observed D_{X_O} and vanished D_{X_V} features are presented as triangles inside of blue and brown boxes respectively. All associations between points D_{X_O} and D_{Y_O} together constitute the marked PP which admits a stochastic kernel $\ell(y|x)$. This indicates that the point x may have any point $y \in D_Y$ as its mark, but intuitively some marks should be more likely than others. In Figure 3 (c) and (d) we give examples of these different matchings, which are indicated by connected pairs inside of the blue box. Finally, the unexpected features in D_{Y_U} are presented as dots inside of red boxes in Figure 3 (c) and (d). The stochastic matching is used as likelihood basically to understand the nature of points on the observed PDs. The posterior intensity and cardinality are given in the theorem below, whose proof is delegated to Section 1.1 in the supplementary materials.

Theorem 3.1. *For a random PD, denote the prior intensity and cardinality by λ_{D_X} and $\rho_{\mathcal{D}_X}$, respectively. Suppose $\alpha(x)$ is the probability of observing a prior feature, and $\mathcal{D}_{X_O}$ and $\mathcal{D}_{X_V}$ are two instances of observed and vanished features in the prior respectively. If $\ell(y|x)$ is the stochastic kernel that links $\mathcal{D}_{Y_O}$ with $\mathcal{D}_{X_O}$, and $\lambda_{\mathcal{D}_{Y_U}}$ and $\rho_{\mathcal{D}_{Y_U}}$ are the intensity and cardinality of $\mathcal{D}_{Y_U}$ respectively, then for a set of independent samples of PDs $D_{Y_{1:m}} = \{D_{Y_1}, \cdots, D_{Y_m}\}$ from $\mathcal{D}_Y$ with cardinalities $K_1, \cdots, K_m$, we have the*

following posterior intensity and cardinality:

$$\lambda_{\mathcal{D}_X|D_{Y_{1:m}}}(x) = \frac{1}{m}\sum_{i=1}^{m}\left[(1-\alpha(x))\lambda_{\mathcal{D}_X}(x)B(\emptyset) + \sum_{y\in D_{Y_i}}\frac{\alpha(x)\ell(y|x)\lambda_{\mathcal{D}_X}(x)B(y)}{\lambda_{\mathcal{D}_{Y_U}}(y)}\right], \quad (3.1)$$

and

$$\rho_{\mathcal{D}_X|D_{Y_{1:m}}}(n) = \frac{1}{m}\sum_{i=1}^{m}\frac{\rho_{\mathcal{D}_X}(n)\left(\sum_{k=0}^{K_i}(K_i-k)!\,P_k^n\,\rho_{\mathcal{D}_{Y_U}}(K_i-k)\;(\lambda_{\mathcal{D}_X}[1-\alpha])^{n-k}\;e_{K_i,k}(D_{Y_i})\right)}{\langle\rho_{\mathcal{D}_X},\Gamma^{0,0}_{D_{Y_i}}\rangle}, \quad (3.2)$$

where

$$B(\emptyset) = \frac{\langle\rho_{\mathcal{D}_X},\Gamma^{0,1}_{D_{Y_i}}\rangle}{\langle\rho_{\mathcal{D}_X},\Gamma^{0,0}_{D_{Y_i}}\rangle}, \quad B(y) = \frac{\langle\rho_{\mathcal{D}_X},\Gamma^{1,1}_{D_{Y_i}\setminus y}\rangle}{\langle\rho_{\mathcal{D}_X},\Gamma^{0,0}_{D_{Y_i}}\rangle}, \quad e_{K_i,k}(D_{Y_i}) = \sum_{\substack{S_{Y_i}\subseteq D_{Y_i}\\|S_{Y_i}|=k}}\prod_{y\in S_{Y_i}}\frac{\lambda_{\mathcal{D}_X}[\alpha\ell(y|x)]}{\lambda_{\mathcal{D}_{Y_U}}(y)},$$

$$\Gamma^{a,b}_{D_{Y_i}}(\tau) = \left(\sum_{k=0}^{\min\{K_i-a,\tau\}}(K_i-k-a)!\,P^{\tau}_{k+b}\,\rho_{\mathcal{D}_{Y_U}}(K_i-k-a)(\lambda_{\mathcal{D}_X}[1-\alpha])^{\tau-k-b}e_{K_i-a,k}(D_{Y_i})\right), \quad (3.3)$$

$f[\zeta] = \int_{\mathcal{X}}\zeta(x)f(x)dx$ *is a linear functional,* P_i^n *is the permutation coefficient, and the sum in* $\Gamma^{0,0}_{D_{Y_i}}(n)$ *of* (3.2) *goes from* 0 *to* K_i.

In the posterior intensity expression given in (3.1), the two terms reflect the decomposition of the prior intensity. Due to the arbitrary cardinality distribution assumption for i.i.d. cluster point processes, the two terms are also weighted by two factors $B(\emptyset)$ and $B(y)$ respectively. The first term is for the vanished features $\mathcal{D}_{X_V}$, where the intensity is weighted by $1-\alpha(x)$ and $B(\emptyset)$. The factor $B(\emptyset)$ is encountered since there is no $y \in D_{Y_i}$ to represent the vanished features $\mathcal{D}_{X_V}$. The second term in (3.1) corresponds to the observed part $\mathcal{D}_{X_O}$ and is weighted by $\alpha(x)$ and $B(y)$. The factor $B(y)$ depends on specific $y \in D_{Y_i}$ to account for the associations between the features in $\mathcal{D}_{X_O}$ and those in D_{Y_i}. To be more precise, if $x \in D_X$ is observed, it can be associated with any of the $y \in D_{Y_i}$ and the remaining points of D_{Y_i}, defined as $D_{Y_i} \setminus y$, are considered to either be observed from the rest of the features in $\mathcal{D}_X$ or originated as unexpected features $\mathcal{D}_{Y_U}$.

The posterior cardinality is given in (3.2). The associated likelihood is given as the sum from $k = 0$ to K_i, where K_i is the number of features in D_{Y_i}. This provides the likelihood of each observed PD D_{Y_i} given that there are n points in $\mathcal{D}_X$. In particular, for $k = 0$, the cardinality term for the unexpected feature reduces to $\rho_{\mathcal{D}_{Y_U}}(K_i)$ and the intensity term for the vanished feature reduces to $(\lambda_{\mathcal{D}_X}[1-\alpha])^n$. This implies that

if the observed PD consists only of unexpected features then all of the points in the prior are most likely to have vanished. As the value of k increases, contributions from the unexpected features and vanished features decrease, indicating the presence of more associations between prior and observed features through the marked point process $(\mathcal{D}_{X_O}, \mathcal{D}_{Y_O})$.

3.1 Closed Form of Posterior Estimation

Next, we present a closed form solution to the posterior intensity and cardinality equation of Theorem 3.1 by considering a Gaussian mixture density for the prior intensity and a binomial distribution for the prior cardinality. Below we specify the necessary components of Theorem 3.1 to derive these closed forms.

(M1) The expressions for the prior intensity $\lambda_{\mathcal{D}_X}$ and cardinality $\rho_{\mathcal{D}_X}$ are:

$$\lambda_{\mathcal{D}_X}(x) = \sum_{l=1}^{N} c_l^{\mathcal{D}_X} \mathcal{N}^*(x; \mu_l^{\mathcal{D}_X}, \sigma_l^{\mathcal{D}_X}\mathbf{I}), \text{ and } \rho_{\mathcal{D}_X}(n) = \binom{N_0}{n} \rho_x^n (1-\rho_x)^{N_o - n}, \quad (3.4)$$

where N is the number of components of the Gaussian mixture, $\mu_i^{\mathcal{D}_X}$ is the mean which is a 2×1 vector of birth and persistence coordinates, and $\sigma_i^{\mathcal{D}_X}\mathbf{I}$ is the 2×2 covariance matrix of i-th component. Since PDs are modeled as point processes on the space $\mathbb{W}$ not on $\mathbb{R}^2$, the Gaussian densities are restricted to $\mathbb{W}$ as $\mathcal{N}^*(z; \upsilon, \sigma\mathbf{I}) := \mathcal{N}(z; \upsilon, \sigma\mathbf{I})\mathbb{1}_{\mathbb{W}}(z)$, with mean v and covariance matrix $\sigma\mathbf{I}$, and $\mathbb{1}_{\mathbb{W}}$ is the indicator function of $\mathbb{W}$. $N_0 \in \mathbb{N}$ is the maximum number of points in the prior PP and $\rho_x \in [0, 1]$ is the probability of one point to fall in the space $\mathbb{W}$.

(M2) The likelihood function $\ell(y|x)$, which is the stochastic kernel of the marked i.i.d. cluster PP $(\mathcal{D}_{X_O}, \mathcal{D}_{Y_O})$, takes the form

$$\ell(y|x) = \mathcal{N}^*(y; x, \sigma^{\mathcal{D}_{Y_O}}\mathbf{I}), \quad (3.5)$$

where $\sigma^{\mathcal{D}_{Y_O}}$ is the covariance coefficient that quantifies the level of confidence in the observations.

(M3) The i.i.d. cluster PP $\mathcal{D}_{Y_U}$, consisting of the unexpected features in the observation, has intensity $\lambda_{\mathcal{D}_{Y_U}}$ and cardinality $\rho_{\mathcal{D}_{Y_U}}$. The intensity for $\mathcal{D}_{Y_U}$ takes the form

$$\lambda_{\mathcal{D}_{Y_U}}(y_{\text{birth}}, y_{\text{pers}}) = \mu^2_{\mathcal{D}_{Y_U}} \exp(-\mu_{\mathcal{D}_{Y_U}}(y_{\text{birth}} + y_{\text{pers}})). \quad (3.6)$$

$\mu_{\mathcal{D}_{Y_U}}$ controls the rate of decay away from the origin. This distribution for $\lambda_{\mathcal{D}_{Y_U}}$ considers points closer to the origin more likely to be unexpected features. Points close to the origin in PDs are often created either from the spacing between the point clouds due to sampling or from the presence of noise in the data. Typically points with higher persistence or higher birth represent significant topological signatures, so for our analysis we count them as less likely to be unexpected. The cardinality distribution is

$$\rho_{\mathcal{D}_{Y_U}}(n) = \binom{M_0}{n} \rho_y^n (1-\rho_y)^{M_0 - n}, \quad (3.7)$$

where $M_0 \in \mathbb{N}$ is the maximum number of points in the PP $\mathcal{D}_{Y_U}$ and $\rho_y \in [0,1]$ is the probability of one point to fall in the space $\mathbb{W}$.

Proposition 3.1. *Suppose that $\lambda_{\mathcal{D}_X}, \rho_{\mathcal{D}_X}, \ell(y|x)$, $\lambda_{\mathcal{D}_{Y_U}}$, and $\rho_{\mathcal{D}_{Y_U}}$ satisfy the assumptions (M1)–(M3), and α is fixed. Then the posterior intensity and cardinality of Theorem 3.1 are given by:*

$$\lambda_{\mathcal{D}_X|D_{Y_{1:m}}}(x) = \frac{1}{m}\sum_{i=1}^{m}\left[(1-\alpha)\lambda_{\mathcal{D}_X}(x)B(\emptyset) + \sum_{y\in D_{Y_i}}\sum_{l=1}^{N} C_l^{x|y}\mathcal{N}^*(x;\mu_l^{x|y},\sigma_l^{x|y}\mathbf{I})\right] \quad (3.8)$$

and

$$\rho_{\mathcal{D}_X|D_{Y_{1:m}}}(n) = \frac{1}{m}\sum_{i=1}^{m}\frac{\rho_{\mathcal{D}_X}(n)\Gamma^{0,0}_{D_{Y_i}}(n)}{\langle \rho_{\mathcal{D}_X}, \Gamma^{0,0}_{D_{Y_i}}\rangle}, \quad (3.9)$$

where $\Gamma^{a,b}_{D_{Y_i}}(\tau)$, $B(\emptyset)$, and $B(y)$ are as in Theorem 3.1 with

$$e_{K_i,k}(D_{Y_i}) = \sum_{S_{Y_i}\subseteq D_{Y_i}, |S_{Y_i}|=k}\prod_{y\in S_{Y_i}}\frac{\alpha\langle c^{\mathcal{D}_X}, q(y)\rangle}{\lambda_{\mathcal{D}_{Y_U}}(y)};\ \ q_l(y) = \mathcal{N}(y;\mu_l^{\mathcal{D}_X},(\sigma^{\mathcal{D}_{Y_O}}+\sigma_l^{\mathcal{D}_X})\mathbf{I});$$

$$C_l^{x|y} = \frac{B(y)\ \alpha c_l^{\mathcal{D}_X} q_l(y)}{\lambda_{D_{Y_U}}(y)};\ \ \mu_l^{x|y} = \frac{\sigma_l^{\mathcal{D}_X} y + \sigma^{\mathcal{D}_{Y_O}}\mu_l^{\mathcal{D}_X}}{\sigma_l^{\mathcal{D}_X}+\sigma^{\mathcal{D}_{Y_O}}};\ \ \textit{and}\ \ \sigma_l^{x|y} = \frac{\sigma^{\mathcal{D}_{Y_O}}\ \sigma_l^{\mathcal{D}_X}}{\sigma_l^{\mathcal{D}_X}+\sigma^{\mathcal{D}_{Y_O}}}.$$

We present the proof in Section 2 of the supplementary materials. One can see that the intensity estimation in (3.8) is in the form of a Gaussian mixture, and hence it is obtained from a conjugate family of priors. However, we do not observe a similar property for the cardinality estimation. A detailed example of these estimations is provided in Section 3.2. The cardinality distribution in (3.9) is computed for infinitely many values of n, which is unattainable. Hence, for the practical application, we must truncate n at some N_{max} such that N_{max} is sufficiently larger than the number of points in the prior PP. Without loss of generality, we can choose $N_{max} = N_0$.

3.2 Sensitivity Analysis

We present the following example to (i) illustrate the estimation of the posterior using (3.8) and (3.9), (ii) examine the effects of the choice of prior intensity and cardinality on the posterior distributions, and (iii) examine the effects of likelihood function and unexpected features parameters on the posterior distributions. To reproduce these results, the interested reader may download our R-package BayesTDA. We consider point clouds generated from a polar curve that contains two inner loops (see Figure 4 (a)) and focus on 1-dimensional features in their corresponding PDs as they are the important homological features of this shape.

Choice of Priors We commence by defining an i.i.d cluster PP with three types of prior intensities: (i) informative, (ii) weakly informative, and (iii) uninformative, and two

Parameters for (M1)				
Prior	$\mu_i^{\mathcal{D}_X}$	$\sigma_i^{\mathcal{D}_X}$	$c_i^{\mathcal{D}_X}$	N_0
Informative	$(0.2, 0.55)$ $(0.17, 0.35)$	0.0018 0.0018	2 2	15
Weakly Informative	$(0.2, 0.55)$ $(0.17, 0.35)$	0.005 0.003	2 2	15
Unimodal Uninformative	$(0.5, 0.5)$	0.5	1	15

Table 1: List of parameters for (M1). We take into account three types of prior intensities: (i) informative, (ii) weakly informative, and (iii) uninformative, and two types of prior cardinalities: (i) informative and (ii) uninformative.

types of prior cardinalities: (i) informative and (ii) uninformative. The prior intensities are considered to be Gaussian mixtures as discussed in (M1). We present all intensity maps on a scale from 0 to 1 throughout this example to ensure uniformity. Due to the symmetric nature of the polar curve, in a noiseless scenario, the corresponding PD consists of two points each with multiplicity two. Hence we use two Gaussian components centered at the two points with a very small variance and weights $c_i^{\mathcal{D}_X} = 2$ for the informative intensity (II) (see Figures 4, 6–7 (b)). The weakly informative intensity (WII) also has two Gaussian components centered at the same points as II with a slightly higher variance ((see Figures 4, 6–7 (c))). The informative cardinality (IC) is determined by using a discrete distribution with the highest probability at cardinality 4 (see Figures 4, 6–7 (e)). On the other hand for the uninformative intensity (UI), we use one Gaussian component centered at an arbitrary point with higher variance than the informative cases as shown in Figures 4, 6–7 (d). Similarly, the uninformative cardinality (UC) follows a discrete uniform distribution (see Figures 4, 6–7 (i)). We present the list of parameters used to define the prior PP in Table 1. We examine the cases below.

The observed PDs are generated from point clouds sampled uniformly from the polar curve and perturbed by varying levels of Gaussian noise with variances $0.001I_2$ (Figure 4 (a)), $0.005I_2$ (Figure 6 (a)), and $0.01I_2$ (Figure 7 (a)) which are considered in Case-1, Case-2, and Case-3 respectively. Consequently, their PDs exhibit distinctive characteristics such as four prominent features with high persistence and very few spurious features, four prominent features with medium persistence and several spurious features, and three prominent features with medium persistence and many spurious features.

Case-1 The point cloud considered here is shown in Figure 4 (a). The 1-dimensional features in the corresponding PD are presented as black triangles overlaid on the posterior intensity plots. We examine the posterior intensity and cardinality for six different combinations of priors – (II, IC), (WII, IC), (UI, IC), (II, UC), (WII, UC), and (UI, UC). As the PD consists of a very low number of spurious features, we observe that the posterior computed from any combination of the six predicts the existence and position of all 1-dimensional features accurately (Figure 4 (f)–(h) and (j)–(l)). The uninformative prior cardinality also produces very low variance in the posterior cardinality estimation.

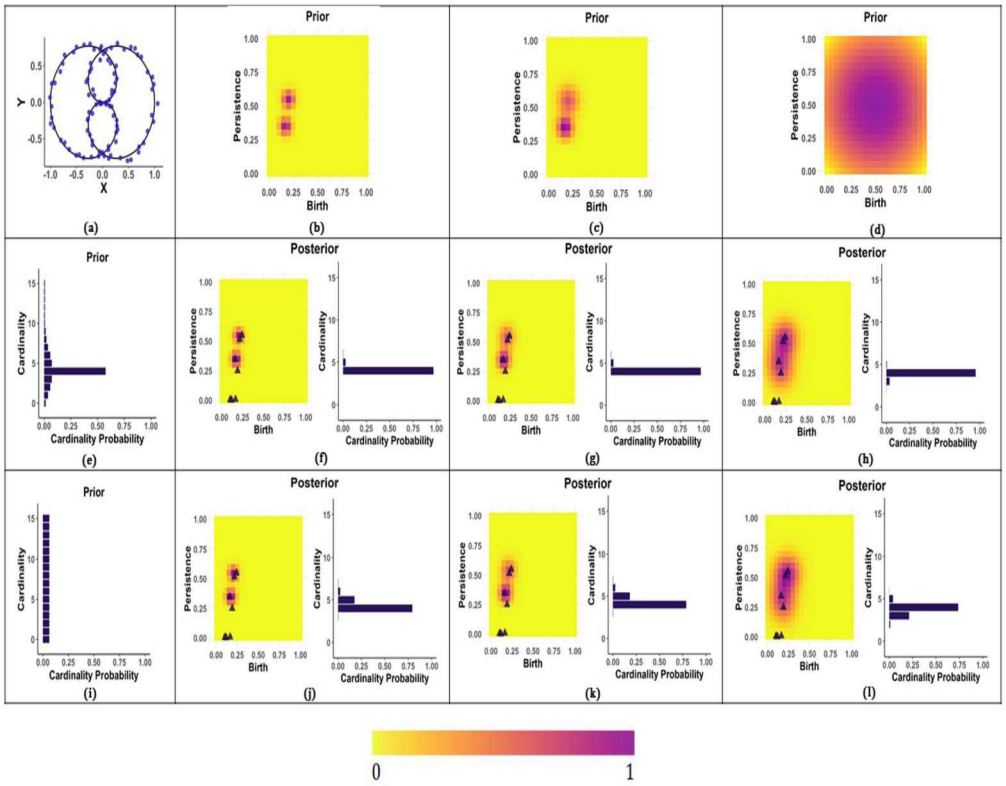


Figure 4: Posterior intensities and cardinalities obtained for Case-1 by using Proposition 3.1. The 1-dimensional features in the corresponding PD are presented as black triangles overlaid on the posterior intensity plots. The color map represents scaled intensities.

Furthermore, for this case we present a comparison between the cardinality statistics given by using i.i.d. cluster point process characterization of the PD presented herein and a Poisson point process framework presented in Maroulas et al. (2020) that estimates the number of homological features by integrating the estimated posterior intensity. As discussed earlier, the Poisson PP framework approximates the cardinality as a Poisson distribution, and consequently this estimation produces higher variability as the number of points increases. However, the i.i.d. cluster PP characterization leads to accurate estimation of the cardinality with tighter variance (see Figure 5).

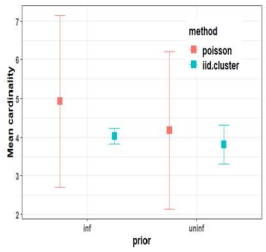


Figure 5: Cardinality statistics for the posterior obtained by using the parameters in Case-1 for Poisson and i.i.d. cluster point process.

Cases		Parameters for (M2)	Parameters for (M3)		
		$\sigma^{\mathcal{D}_{Y_O}}$	$\mu^{\mathcal{D}_{Y_U}}$	ρ_y	M_0
Case-1	(e),(f),(h),(i)	0.01	20	0.5	15
Case-2	(e),(f),(h),(i) (k) (l)	0.01 0.001 0.001	20 25 16	0.5	15
Case-3	(e),(f),(h),(i) (k)–(l)	0.01 0.001	20	0.5 0.6	15

Table 2: List of parameters for (M2) and (M3). For Cases-1, 2, and 3, we consider point clouds sampled from the polar curve and perturbed by Gaussian noise having variances $0.001I_2$, $0.005I_2$, and $0.01I_2$ respectively.

Case-2 We consider all of the priors as in Case-1. The point cloud used for this case (Figure 6 (a)) is more perturbed around the polar curve than Case-1 (Gaussian noise with variance $0.005I_2$). The associated PD, presented as black triangles overlaid on the posterior intensity plots, exhibits more spurious features. The parameters used for this case are listed in Table 2. First, we estimate the posterior intensity and cardinality for all six combinations using the same parameters as in Case-1, and the results are presented in Figure 6 (f)–(h) and (j)–(l). For the combinations (II, IC), (WII, IC), and (UI, IC) of priors, the posterior intensity and cardinality can accurately estimate the holes with different variance levels. As the WII equipped with variance higher than II, the variance of posterior cardinality obtained from (WII, IC) is slightly higher than that of (II, IC). However, due to the presence of several spurious features the three combinations, (II, UC), (WII, UC), and (UI, UC), slightly overestimate the cardinality. Next, to illustrate the effect of observed data on the posterior, we adjust two parameters, the variance of the likelihood $\sigma_{D_{Y_O}}$ and the decay parameter of the unexpected features, $\mu_{D_{Y_U}}$. Recall that the intensity density of the PP $\mathcal{D}_{Y_U}$, consisting of the unexpected features in the observation, is exponential (Equation (3.6)), where $\mu_{D_{Y_U}}$ controls the rate of decay away from the origin. We present the updated posteriors from the three combinations of priors (II, UC), (WII, UC), and (UI, UC). By decreasing the variance of the likelihood $\sigma_{D_{Y_O}}$, the posterior intensities rely more on the observed features in the PD (see Figure 6 (n)–(p)). On the other hand, by adjusting the decay parameter, we enable our model to recognize the presence of several spurious features in PD. This improves the estimation of posterior cardinality, which is evident in Figure 6 (n)–(p).

Case-3 In this case we consider the point cloud (Figure 7 (a)), which is very noisy (Gaussian noise with variance $0.01I_2$). Due to the noise level, we encounter only three points with medium prominent persistence, and there are many spurious features. All the priors are the same as in Case-1 and Case-2. The associated PD is presented as black triangles overlaid on the posterior intensity plots. The parameters used for this case are listed in Table 2. First, we estimate the posterior intensity and cardinality for all six combinations using the same parameters as in Case-1, and the results are presented in Figure 7 (f)–(h) and (j)–(l). For the combinations (II, IC), (WII, IC), and (UI, IC) of priors, the posterior intensity and cardinality can accurately estimate the position and number of 1-dimensional features with different variance levels. The

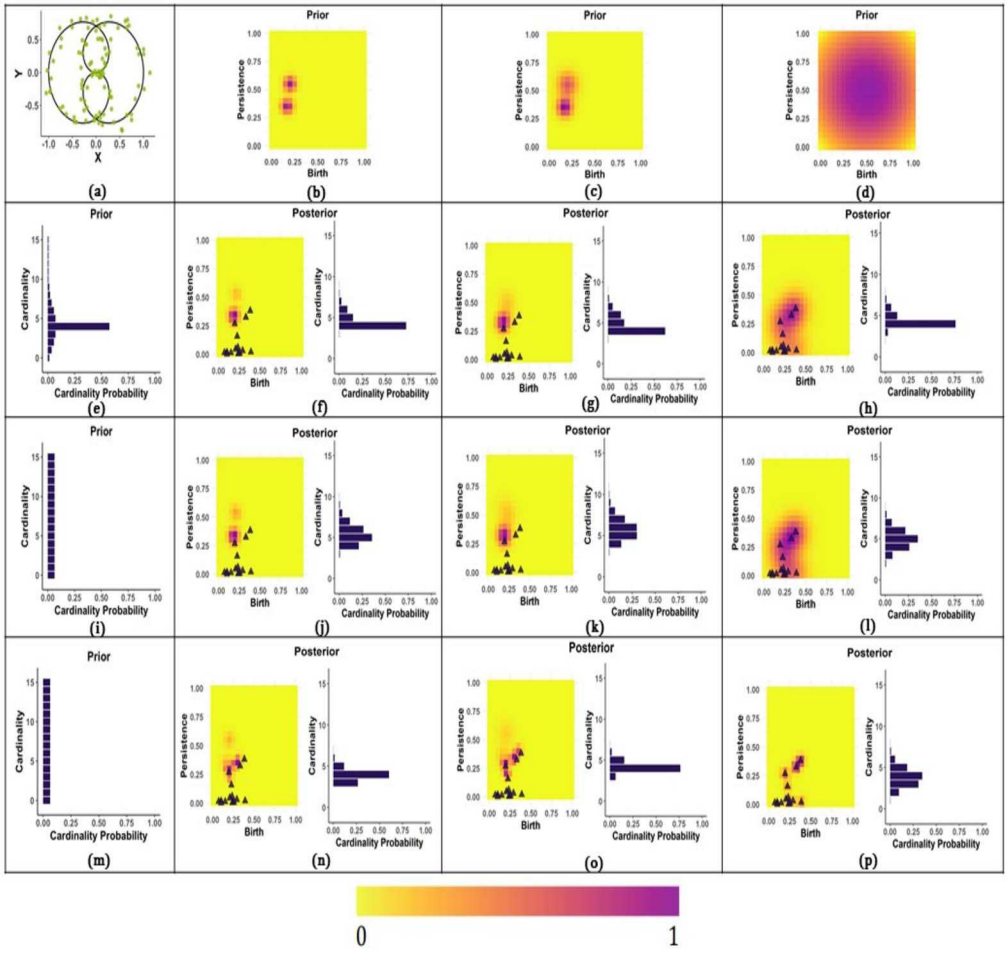


Figure 6: Posterior intensities and cardinalities obtained for Case-2 by using Proposition 3.1. The 1-dimensional features in the corresponding PD are presented as black triangles overlaid on the posterior intensity plots. The color map represents scaled intensities.

posterior estimates from (WII, IC) show higher variability than that of (II, IC). Due to the presence of several spurious features, the other three combinations (II, UC), (WII, UC), and (UI, UC) overestimate the cardinality distribution. Also, in the (UI, UC) case the posterior intensity estimates the location of the hole with higher variance and is skewed towards the noise features. Next, to illustrate the effect of the observed features on the posterior, we adjust two parameters, the variance of the likelihood $\sigma_{D_{Y_O}}$ and the unexpected feature cardinality parameter ρ_y in the posterior estimation for the three combinations (II, UC), (WII, UC), and (UI, UC). By decreasing $\sigma_{D_{Y_O}}$ we notice that the posterior intensities rely more on the observed features in the PD (see Figure 7 (n)–(p)). On the other hand by increasing ρ_y the model is able to identify that there are

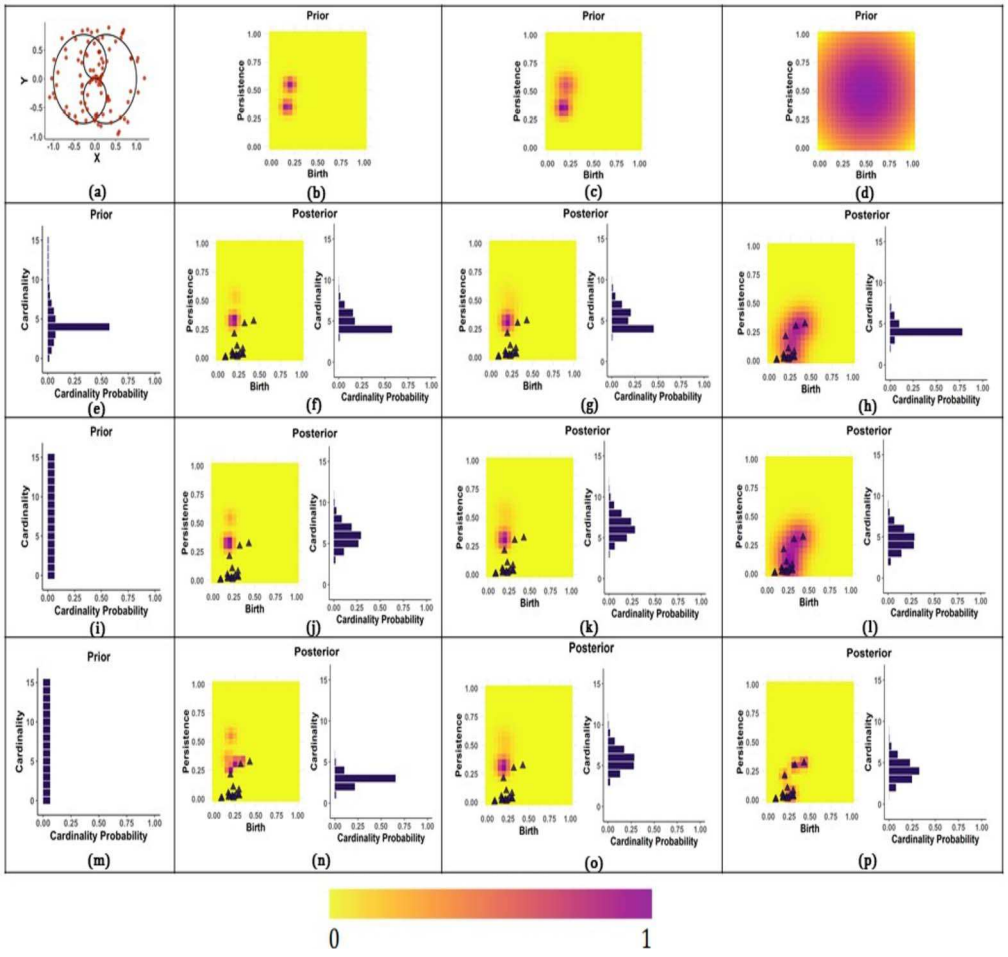


Figure 7: Posterior intensities and cardinalities obtained for Case-3 by using Proposition 3.1. The 1-dimensional features in the corresponding PD are presented as black triangles overlaid on the posterior intensity plots. The color map represents scaled intensities.

more spurious features in this PD than that of Case-1 and Case-2. This improves the estimation of posterior cardinality (see Figure 7 (n)–(p)).

4 Classification of Actin Filament Networks

In this section, we classify 150 simulated actin filament networks in plant cells. Such filaments are key in the study of intracellular transportation in plant cells, as these bundles and networks make up the actin cytoskeleton, which determines the structure of the cell and enables cellular motion. In particular, three different classes of networks

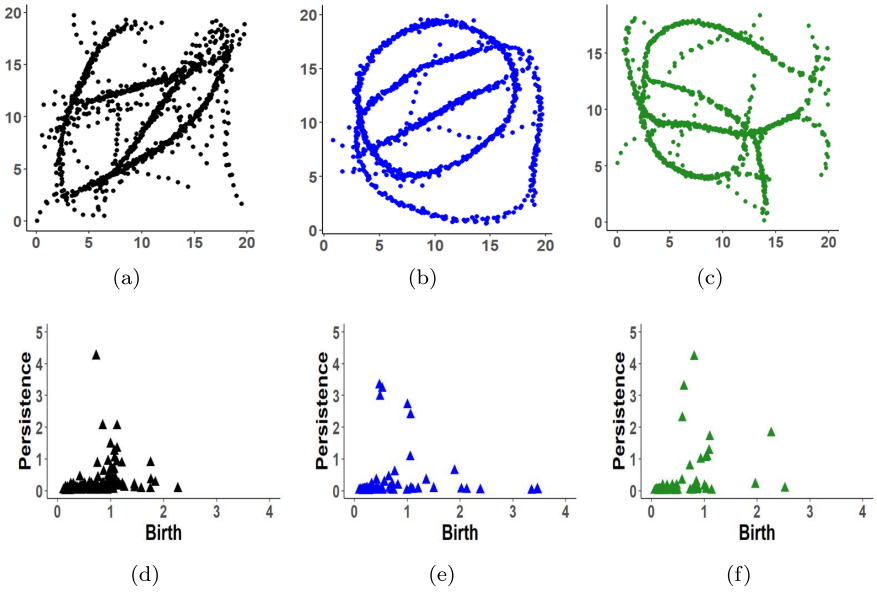


Figure 8: (a)–(c) are examples of PDs generated from networks in $\mathcal{C}_1, \mathcal{C}_2$, and $\mathcal{C}_3$ respectively. (d)–(f) are their corresponding PDs.

with different numbers of cross linking proteins were considered. The number of cross linking proteins in the network affects the shape it. The networks were created using the AFINES (Active Filament Network Simulation) stochastic simulation framework introduced in Freedman et al. (2018, 2017), which models the assembly of the actin cytoskeleton. In the data set the location of beads that represents segments of individual filaments are known. The value of each parameter in the simulation process is chosen to mimic real actin filaments. Hence our Bayesian topological learning method herein can be implemented directly to real actin filament network data sets.

Higher numbers of cross-linking proteins produce local geometric signatures (Tang et al. (2014)). However, the differences are not always notable due to the presence of noise in the data, which is a routine scenario for real experiments. To that end, we adopt a data-driven scheme for classification using an uninformative flat prior. We learn the networks in training sets by means of their respective PDs as they distill salient information about the network patterns with respect to connectedness and empty space (holes), i.e. we can differentiate between filament networks by examining their homological features.

The three classes generated with the cross-linking proteins numbers of $N = 825, 1650,$ and 3300 are denoted as $\mathcal{C}_1, \mathcal{C}_2$, and $\mathcal{C}_3$, respectively (see Figure 8 (a)–(c) for examples). From the viewpoint of topology, class $\mathcal{C}_2$ and class $\mathcal{C}_3$ contain more prominent holes than class $\mathcal{C}_1$. Also, their respective PDs have different cardinalities. Hence, this topological aspect yields an important contrast between these three classes. To capture these differences we employ the following Bayes factor classification approach

Parameters for (M1)				
$\mu_i^{\mathcal{D}_X}$	$\sigma_i^{\mathcal{D}_X}$	$c_i^{\mathcal{D}_X}$	N_0	$\rho_{\mathcal{D}_X}$
$(1,2)$	6	1	25	$24/25$
Parameters for (M2)		**Parameters for (M3)**		
$\sigma^{\mathcal{D}_{Y_O}}$	$\mu^{\mathcal{D}_{Y_U}}$	M_0	ρ_y	α
0.01	1	25	$2/25$	0.95

Table 3: List of parameters used for the classification.

by relying on the closed form estimation of posterior distributions discussed in Section 3.1. A PD D that needs to be classified is a sample from an i.i.d. cluster point process $\mathcal{D}$ with intensity $\lambda_{\mathcal{D}}$ and cardinality $\rho_{\mathcal{D}}$ and its probability density has the form $p_{\mathcal{D}}(D) = \rho_{\mathcal{D}}(|D|)\prod_{d\in D}\lambda_{\mathcal{D}}(d)$. For a training set $Q_{Y^k} := D_{Y^k_{1:n}}$ for $k = 1, \cdots, K$ from K classes of random diagrams $\mathcal{D}_{Y^k}$, we obtain the posterior intensities from the Bayesian framework using Proposition 3.1. The posterior probability density of D given the training set Q_{Y^k} is given by

$$p_{\mathcal{D}|\mathcal{D}_{Y^k}}(D|Q_{Y^k}) = \rho_{\mathcal{D}|\mathcal{D}_{Y^k}}(|D|)\prod_{d\in D}\lambda_{D|Q_{Y^k}}(d), \tag{4.1}$$

and consequently, the Bayes factor is obtained by the ratio $BF^{ij}(Q_{Y^i}, Q_{Y^j}) = \frac{\rho_{D|\mathcal{D}_{Y^i}}(D|Q_{Y^i})}{\rho_{D|\mathcal{D}_{Y^j}}(D|Q_{Y^j})}$ for a class $i, j = 1, \cdots, K$ such that $i \neq j$. For every pair (i, j), if $BF^{ij}(Q_{Y^i}, Q_{Y^j}) > c$, we assign one vote to class Q_{Y^i}, or otherwise for $BF^{ij}(Q_{Y^i}, Q_{Y^j}) < c$. The final assignment of the class of D is obtained by a majority voting scheme.

PDs with 1-dimensional features (see Figure 8 (d)–(f) for an example of each class) were created for each actin network through Rips filtration as discussed in Section 2.1, which were then used as input for the Bayes factor classification scheme of (4.1). The number of 1-dimensional features in the dataset is large and the posterior estimation for this dataset is not computationally attainable. To mitigate this issue, we subsample the dataset to reduce the size of it. Precisely, our subsampled dataset consists of 25 points from each of the PDs obtained from the 150 simulated filament networks. We found that taking more than 25 points from each of the PDs did not improve the classification, and typically led to a very expensive computational scheme. Table 3 summarizes the choices of parameters for the model. The intensity density of $\mathcal{D}_{Y_U}$ used for the classification is presented in Figure 9.

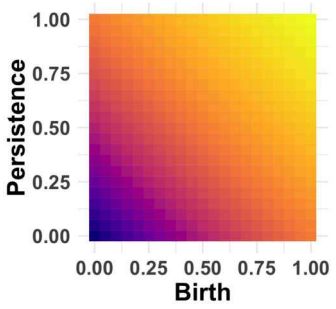


Figure 9: The intensity density for the unexpected feature PP used in classifying the filament networks.

One intuitive interpretation of the unexpected features is that they represent the presence of noise in the dataset, consequently they often have very short persistence. On the other hand, the dataset of filament networks routinely consists of several incomplete

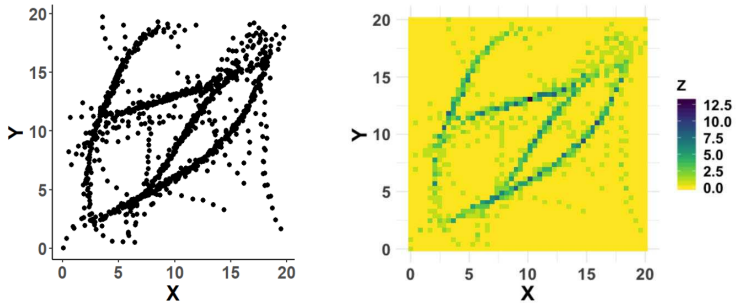
loops, which imply that points with late birth and short persistence are expected from the underlying topology. Since we use 10-fold cross validation to estimate the model's accuracy, the posterior is calculated using the training set for each fold and each class. Then for each instance, we assign the class by using the majority voting scheme. We compute the resulting area under the receiver operating characteristic (ROC) curves (AUCs) and the results are listed in Table 4. The AUC across 10-folds was 0.925. Further details on the classification problem are given in the supplementary materials.

4.1 Comparison with Other Methods

We compared our method with several other machine learning algorithms to benchmark against them. We mainly pursued two avenues – (i) features selected using TDA methodology, and (ii) features selected using non-TDA methodology. The TDA methodologies are persistence images (PIs) (Adams et al., 2017), persistence landscapes (PLs) (Bubenik, 2015), and Euler characteristic curves (ECCs) (Richardson and Werman, 2014). These summaries have been widely implemented as they are amenable to the existing machine learning methodologies. The main theme of these summaries is the extraction of a pertinent feature vector and implement a classifier trained using machine learning algorithms. Here we input these topological summaries as features for three different optimized classification algorithms: random forest (RF), support vector machine (SVM), and neural network (NN).

We considered a vector of 2500 values at which the PLs of order 1, 2, and 3 are evaluated, and found that the third order PL to be the most efficient summary for this classification task. In order to compute the PIs, we discretize the domain space into a 50×50 grid with a spread of 0.1. The linear ramp function is used to produce weights for computing PIs. We explore the classification problem using PIs with and without incorporating the linear weights and found that the PIs without any weights provide better accuracy than those with weights. This is justified as the linear ramp function assigns more weights to the higher persistence points leaving the local features to be insignificant. Another topological summaries we implemented are ECCs from 0 and 1-dimensional persistence diagrams where the filtration were constructed from the range of birth and persistence values. We optimally tune the parameters of SVM using a grid search. Precisely, the parameter γ of the radial basis kernel, that is the inverse of the standard deviation of the kernel, was optimally selected from a range of 0.1 to 1 with a spread of 0.1. In order to choose the optimal parameters for NN, we performed an extensive grid search for all parameters. However, we found that out of all the parameters, the only two that can potentially improve the classification accuracy are the number of hidden layers and the maximum number of iterations. The optimal performance was achieved for PLs with 20 layers and maximum iterations of 10, for PIs with 3 layers and maximum iterations of 200, and for ECCs with 20 layers and maximum iterations of 10. For the RF algorithm we employ 500 trees.

Additionally, we compare our method with machine learning algorithms where the features are selected using a non-TDA method. As the filament networks pose a very definite spatial structure, we found the most useful method to extract the feature is the Raster images Hijmans (2019). In particular, the raster image represents data by using



(a) (b)

Figure 10: (a) An example filament network from $\mathcal{C}_1$. (b) The network in (a) converted to a raster image.

a grid with a value assigned for each pixel. The assigned value can reflect a wide variety of information. In our analysis, we discretize the domain of a filament network into 2500 grid cells identified by 50 rows and 50 columns, and then count the number the points of each grid cell. This approach not only converts each filament network into a raster image which in turn is used as input to machine learning algorithms but also captures the definite spatial structures such as the presence of empty space and connectedness in a very efficient manner. We present an example in Figure 10. The parameters for the machine learning algorithms are tuned in a similar fashion, i.e., the parameters are optimally tuned using a grid search. The optimal performance for NN was achieved with 5 layers and maximum iterations of 200. The results of this comparison are in Table 4, which showcases that our method outperforms the other methods.

Method	**AUC**	**Method**	**AUC**
Bayesian Framework	0.925	Random Forest ECC	0.81
Random Forest PI	0.90	SVM ECC	0.78
SVM PI	0.85	NN ECC	0.77
Neural Net PI	0.88	Random Forest Raster	0.69
Random Forest PL	0.82	SVM Raster	0.77
SVM PL	0.72	Neural Net Raster	0.6
Neural Net PL	0.79		

Table 4: Comparison of methods for filament networks.

5 Discussion

This paper has proposed a generalized Bayesian framework for PDs by modeling them as i.i.d. cluster point processes. Our framework provides a probabilistic descriptor of the diagrams by simultaneously estimating the cardinality and spatial distributions of points on a PD. In this work we focus on developing Bayesian model for PDs only.

A fusion of different topological summaries engaged within a Bayesian perspective is a worthwhile future direction.

It is noteworthy that our Bayesian model directly employs PDs, which are topological summaries of data, for defining a substitution likelihood rather than using the entire point cloud. This deviates from a strict Bayesian model, as we consider the statistics of PDs rather than the underlying datasets used to create them; however, our paradigm incorporates prior knowledge and observed data summaries to create posterior distributions, analogous to the notion of substitution likelihood in Jeffreys (1961). The general relationship between the likelihood models related to point cloud data and those of their corresponding persistence diagrams remains an important open problem. We demonstrate that a valid update of the prior distribution on persistence diagrams to the posterior can be made by substitution of the likelihood through a topological summary of the data rather than a traditional likelihood function. Indeed, the idea of utilizing topological summaries of point clouds in place of the actual point clouds proves to be a powerful tool with applications in wide-ranging fields. This process incorporates topological descriptors of point clouds, which simultaneously decipher essential shape peculiarities and avoid unnecessarily complex geometric features.

We derive closed forms of the posterior for realistic implementation, using Gaussian mixtures for the prior intensity and binomials for the prior cardinality. A detailed example showcases the posterior intensities and cardinalities for various interesting instances created by varying parameters within the model. This example exhibits our method's ability to recover the underlying PD. Thus, the Bayesian inference developed here opens up new avenues for machine learning algorithms and data analysis techniques to be applied *directly* to the space of PDs. Indeed, we derive a classification algorithm and successfully apply it to simulated filament networks data, while we compare our method with other TDA and machine learning approaches successfully.

Supplementary Material

Supplementary material for: Bayesian Topological Learning for Classifying the Structure of Biological Networks (DOI: 10.1214/21-BA1270SUPP; .pdf).

References

Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., and Ziegelmeier, L. (2017). "Persistence images: A stable vector representation of persistent homology." *The Journal of Machine Learning Research*, 18(1): 218–252. MR3625712. 2, 20

Banerjee, N. and Park, J. (2015). "Modeling and simulation of biopolymer networks: Classification of the cytoskeleton models according to multiple scales." *Korean Journal of Chemical Engineering*, 32: 1207–1217. 2

Biscio, C. A. and Møller, J. (2019). "The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery

trees and spatial point process applications." *Journal of Computational and Graphical Statistics*, 1537–2715. MR4007749. doi: https://doi.org/10.1080/10618600.2019.1573686. 2

Bobrowski, O., Mukherjee, S., and Taylor, J. E. (2017). "Topological consistency via kernel estimation." *Bernoulli*, 23(1): 288–328. MR3556774. doi: https://doi.org/10.3150/15-BEJ744. 3

Breuer, D., Nowak, J., Ivakov, A., Somssich, M., Persson, S., and Nikoloski, Z. (2017). "System-wide organization of actin cytoskeleton determines organelle transport in hypocotyl plant cells." *Proceedings of the National Academy of Sciences*, 114(28): E5741–E5749. 1

Bubenik, P. (2015). "Statistical topological data analysis using persistence landscapes." *Journal of Machine Learning Research*, 16: 77–102. MR3317230. 2, 20

Carlsson, G. and de Silva, V. (2010). "Zigzag persistence." *Foundations of computational mathematics*, 10(4): 367–405. MR2657946. doi: https://doi.org/10.1007/s10208-010-9066-0. 2

Daley, D. J. and Vere-Jones, D. (1988). *An introduction to the theory of point processes*. New York: Springer-Verlag. MR0950166. 3, 5, 7

Di Fabio, B. and Ferri, M. (2015). "Comparing persistence diagrams through complex vectors." In *International Conference on Image Analysis and Processing*, 294–305. Springer. MR3446500. doi: https://doi.org/10.1007/978-3-319-23231-7_27. 2

Dłotko, P., Ghrist, R., Juda, M., and Mrozek, M. (2012). "Distributed computation of coverage in sensor networks by homological methods." *Applicable Algebra in Engineering, Communication and Computing*, 23(1–2): 29–58. MR2917116. doi: https://doi.org/10.1007/s00200-012-0167-7. 2

Edelsbrunner, H. and Harer, J. L. (2010). *Computational topology: an introduction*. Providence, R.I.: American Mathematical Society. MR2572029. doi: https://doi.org/10.1090/mbk/069. 2

Fasy, B. T., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S., Singh, A., et al. (2014). "Confidence sets for persistence diagrams." *The Annals of Statistics*, 42(6): 2301–2339. MR3269981. doi: https://doi.org/10.1214/14-AOS1252. 3

Freedman, S. L., Banerjee, S., Hocky, G. M., and Dinner, A. R. (2017). "A versatile framework for simulating the dynamic mechanical structure of cytoskeletal networks." *Biophysical journal*, 113(2): 448–460. 1, 18

Freedman, S. L., Hocky, G. M., Banerjee, S., and Dinner, A. R. (2018). "Nonequilibrium phase diagrams for actomyosin networks." *Soft matter*, 14(37): 7740–7747. 18

Guo, W., Manohar, K., Brunton, S. L., and Banerjee, A. G. (2018). "Sparse-TDA: Sparse Realization of Topological Data Analysis for Multi-Way Classification." *IEEE Transactions on Knowledge and Data Engineering*, 30(7): 1403–1408. 2

Higaki, T., Kutsuna, N., Sano, T., Kondo, N., and Hasezawa, S. (2010). "Quantification and cluster analysis of actin cytoskeletal structures in plant cells: role of actin

bundling in stomatal movement during diurnal cycles in Arabidopsis guard cells." *The Plant Journal*, 61(1): 156–165. doi: https://doi.org/10.1111/j.1365-313X.2009.04032.x. 2

Hijmans, R. J. (2019). *raster: Geographic Data Analysis and Modeling*. R package version 3.0-2. URL https://CRAN.R-project.org/package=raster. 20

Humphreys, D. P., McGuirl, M. R., Miyagi, M., and Blumberg, A. J. (2019). "Fast Estimation of Recombination Rates Using Topological Data Analysis." *GENETICS*. doi: https://doi.org/10.1534/genetics.118.301565. 2

Jeffreys, H. (1961). *Theory of Probability*. Clarendon Press. MR0187257. 3, 22

Kerber, M., Morozov, D., and Nigmetov, A. (2017). "Geometry helps to compare persistence diagrams." *Journal of Experimental Algorithmics (JEA)*, 22: 1–4. MR3707737. doi: https://doi.org/10.1145/3064175. 3

Khasawneh, F. A. and Munch, E. (2016). "Chatter detection in turning using persistent homology." *Mechanical Systems and Signal Processing*, 70–71: 527 – 541. 2

Kimura, M., Obayashi, I., Takeichi, Y., Murao, R., and Hiraoka, Y. (2018). "Non-empirical identification of trigger sites in heterogeneous processes using persistent homology." *Scientific reports*, 8(1): 3553. 2

Madison, S. L. and Nebenführ, A. (2013). "Understanding myosin functions in plants: are we there yet?" *Current Opinion in Plant Biology*, 16(6): 710–717. 1

Marchese, A. and Maroulas, V. (2016). "Topological learning for acoustic signal identification." In *2016 19th International Conference on Information Fusion (FUSION)*, 1377–1381. 2

Marchese, A. and Maroulas, V. (2018). "Signal classification with a point process distance on the space of persistence diagrams." *Advances in Data Analysis and Classification*, 12(3): 657–682. MR3736436. doi: https://doi.org/10.1007/s11634-017-0304-z. 2, 3

Maroulas, V., Micucci, C. P., and Nasrin, F. (2021). "Supplementary material for: Bayesian Topological Learning for Classifying the Structure of Biological Networks." *Bayesian Analysis*. doi: https://doi.org/10.1214/21-BA1270SUPP. 4

Maroulas, V., Mike, J. L., and Oballe, C. (2019). "Nonparametric Estimation of Probability Density Functions of Random Persistence Diagrams." *Journal of Machine Learning Research*, 20(151): 1–49. URL http://jmlr.org/papers/v20/18-618.html. MR4030165. 3

Maroulas, V., Nasrin, F., and Oballe, C. (2020). "A Bayesian Framework for Persistent Homology." *SIAM Journal on Mathematics of Data Science*, 2(1): 48–74. MR4060450. doi: https://doi.org/10.1137/19M1268719. 2, 3, 14

Maroulas, V. and Nebenführ, A. (2015). "Tracking Rapid Intracellular movements: a Bayesian random set approach." *The Annals of Applied Statistics*, 9(2): 926–949. MR3371342. doi: https://doi.org/10.1214/15-AOAS819. 2

Mike, J., Sumrall, C. D., Maroulas, V., and Schwartz, F. (2016). "Nonlandmark classification in paleobiology: computational geometry as a tool for species discrimination." *Paleobiology*, 42(4): 696–706. 2

Mileyko, Y., Mukherjee, S., and Harer, J. (2011). "Probability measures on the space of persistence diagrams." *Inverse Problems*, 27(12): 124007. MR2854323. doi: https://doi.org/10.1088/0266-5611/27/12/124007. 3

Mlynarczyk, P. J. and Abel, S. M. (2019). "First passage of molecular motors on networks of cytoskeletal filaments." *Physical Review E*, 99: 022406. doi: https://doi.org/10.1103/PhysRevE.99.022406. 1

Moyal, J. E. (1962). "The general theory of stochastic population processes." *Acta Mathematica*, 108(1): 1–31. MR0148107. doi: https://doi.org/10.1007/BF02545761. 7

Nasrin, F., Oballe, C., Boothe, D. L., and Maroulas, V. (2019). "Bayesian Topological Learning for Brain State Classification." In *Proceedings of 2019 IEEE International Conference on Machine Learning and Applications (ICMLA)*. 2

Nicolau, M. M. P., Levine, A. J., and Carlsson, G. E. (2011). "Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival." *Proceedings of the National Academy of Sciences*, 108(17): 7265–70. 2

Patrangenaru, V., Bubenik, P., Paige, R. L., and Osborne, D. (2018). "Topological Data Analysis for Object Data." arXiv:1804.10255. MR3982197. doi: https://doi.org/10.1007/s13171-018-0137-7. 2

Porter, K. and Day, B. (2016). "From filaments to function: the role of the plant actin cytoskeleton in pathogen perception, signaling and immunity." *Journal of integrative plant biology*, 58(4): 299–311. 1

Reininghaus, J., Huber, S., Bauer, U., and Kwitt, R. (2015). "A stable multi-scale kernel for topological machine learning." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4741–4748. 2

Richardson, E. and Werman, M. (2014). "Efficient classification using the Euler characteristic." *Pattern Recognition Letters*, 49: 99–106. URL https://www.sciencedirect.com/science/article/pii/S0167865514002050. 20

Robinson, A. and Turner, K. (2017). "Hypothesis testing for topological data analysis." *Journal of Applied and Computational Topology*, 1(2): 241–261. MR3975554. doi: https://doi.org/10.1007/s41468-017-0008-7. 3

Sgouralis, I., Nebenführ, A., and Maroulas, V. (2017). "A Bayesian Topological Framework for the Identification and Reconstruction of Subcellular Motion." *SIAM Journal on Imaging Sciences*, 10(2): 871–899. MR3662029. doi: https://doi.org/10.1137/16M1095755. 2

Shimmen, T. and Yokota, E. (2004). "Cytoplasmic streaming in plants." *Current Opinion in Cell Biology*, 16(1): 68–72. 1

Sizemore, A. E., Phillips-Cremins, J. E., Ghrist, R., and Bassett, D. S. (2018). "The importance of the whole: Topological data analysis for the network neuroscientist." *Network Neuroscience*. 2

Staiger, C., Baluska, F., Volkmann, D., and Barlow, P. (2000). *Actin: A Dynamic Framework for Multiple Plant Cell Functions*. Amsterdam: Springer, 1st edition. 1

Tang, H., Laporte, D., and Vavylonisa, D. (2014). "Actin cable distribution and dynamics arising from cross-linking, motor pulling, and filament turnover." *Molecular Biology of the Cell*, 25(19): 3006–3016. 2, 18

Thomas, C., Tholl, S., Moes, D., Dieterle, M., Papuga, J., Moreau, F., and Steinmetz, A. (2009). "Actin bundling in plants." *Cell motility and the cytoskeleton*, 66(11): 940–957. 1

Townsend, J., Micucci, C. P., Hymel, J. H., Maroulas, V., and Vogiatzis, K. D. (2020). "Representation of molecular structures with persistent homology for machine learning applications in chemistry." *Nature Communications*, 11: 3230. 2

Turner, K., Mileyko, Y., Mukherjee, S., and Harer, J. (2014). "Fréchet means for distributions of persistence diagrams." *Discrete and Computational Geometry*, 52(1): 44–70. MR3231030. doi: https://doi.org/10.1007/s00454-014-9604-7. 2

Venkataraman, V., Ramamurthy, K. N., and Turaga, P. (2016). "Persistent homology of attractors for action recognition." In *2016 IEEE International Conference on Image Processing (ICIP)*, 4150–4154. 2

Acknowledgments

We thank the associate editor and two anonymous reviewers for their comments, which helped us to improve our manuscript. We also thank the Abel Research Group for providing the simulated filament network data. The work has been partially supported by the ARO W911NF-21-1-0094, NSF MCB-1715794 and DMS-1821241, and ARL Co-operative Agreement # W911NF-19-2-0328. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes not withstanding any copyright notation herein.