




Topological Data Analysis Approaches to Uncovering the Timing of Ring Structure Onset in Filamentous Networks

Maria-Veronica Ciocanel¹  · Riley Juenemann² · Adriana T. Dawes³ · Scott A. McKinley²

Received: 12 October 2019 / Accepted: 11 December 2020 / Published online: 16 January 2021
© The Author(s) 2021

Abstract

In developmental biology as well as in other biological systems, emerging structure and organization can be captured using time-series data of protein locations. In analyzing this time-dependent data, it is a common challenge not only to determine whether topological features emerge, but also to identify the timing of their formation. For instance, in most cells, actin filaments interact with myosin motor proteins and organize into polymer networks and higher-order structures. Ring channels are examples of such structures that maintain constant diameters over time and play key roles in processes such as cell division, development, and wound healing. Given the limitations in studying interactions of actin with myosin *in vivo*, we generate time-series data of protein polymer interactions in cells using complex agent-based models. Since the data has a filamentous structure, we propose sampling along the actin filaments and analyzing the topological structure of the resulting point cloud at each time. Building on existing tools from persistent homology, we develop a topological data analysis (TDA) method that assesses effective ring generation in this dynamic data. This method connects topological features through time in a path that corresponds to emergence of organization in the data. In this work, we also propose methods for assessing whether the topological features of interest are significant and thus whether they contribute to the formation of an emerging hole (ring channel) in the simulated protein interactions. In particular, we use the MEDYAN simulation platform to show that this technique can distinguish between the actin cytoskeleton organization resulting from distinct motor protein binding parameters.

Keywords Topological data analysis · Actin filament networks · Ring channels · Intracellular transport · Myosin

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11538-020-00847-3>.

✉ Maria-Veronica Ciocanel
ciocanel@math.duke.edu

Extended author information available on the last page of the article

Mathematics Subject Classification 55-04 · 92-00 · 92C37

1 Introduction

Topological data analysis (TDA) has emerged as a new and important set of statistical tools for extracting structural information from high-dimensional data sets (Edelsbrunner et al. 2002; Edelsbrunner and Harer 2008, 2010). In particular, methods from persistent homology are useful in understanding topological invariants such as clusters or loops in data represented as point clouds. Applications of these topological methods in the biological sciences are varied and include quantitative understanding of aggregations such as insect swarms (Topaz et al. 2015; Ulmer et al. 2019), extracting the topology of functional brain networks from fMRI data (Saggar et al. 2018; Stolz et al. 2018), and mapping of unknown spatial environments using biobotic insects (Dirafzoon et al. 2016). While the applications for TDA multiply, there is an associated need for statistical methods that can rigorously compare the behaviors of systems when they are subjected to different environments or controls (Wasserman 2018). A particular challenge for TDA methods is that complex data sets are often corrupted by noisy observations and/or missing data (Chazal et al. 2017; Bobrowski et al. 2017). Moreover, the methods of TDA often exhibit spurious topological features that can vastly outnumber the “real” topological features of interest (Fasy et al. 2014). This last issue is a particular challenge for the biological systems that motivate the work we present here.

1.1 Biological Context: Ring Formation in Filamentous Networks

There are numerous active scientific questions concerning the formation and maintenance of ring-like actin filament structures, which play key roles in developmental and physiological processes (Schwayer et al. 2016). These ring channels are usually composed of actin filaments cross-linked with myosin motor proteins as well as other regulatory binding proteins that control the spatiotemporal organization of the filaments into circular structures in living systems (Schwayer et al. 2016). These actin–myosin rings have been shown to participate in both actively contracting rings, such as those found in cytokinesis and wound healing in organisms ranging from plants to mammals (Robinson and Cooley 1996; Schwayer et al. 2016), as well as in stable ring-like structures that are often used as inter-cellular bridges by germline cells to share nutrients and gene products during development (Robinson and Cooley 1996).

We focus on stable biological ring-like structures that maintain a constant diameter over a long time. Two specific examples of such stable structures occur in developing germline cells in *Drosophila* fruit flies and in *Caenorhabditis* nematode worms. In *Drosophila* fruit fly development, ovarian ring canals connect germ cells and promote transport of cellular components to the developing egg (Robinson et al. 1994; Hudson et al. 2015). The complex assembly and maintenance of these ring canals is not fully understood and there is evidence that various specialized cytoskeletal proteins may regulate their development (Robinson et al. 1994; Ong et al. 2010; Hudson et al. 2015).

Similarly, the actin cytoskeleton forms complex cellular structures in the reproductive system of the worm *C. elegans* (Kelley and Cram 2019). Here, actin filaments interact with non-muscle myosin II motor proteins and other actin-binding proteins to allow for streaming of cytoplasm into enlarging oocytes (Wolke et al. 2007; Osorio et al. 2018). The stable circular structures that emerge are called ring channels and maintain a constant diameter during the development of germ cells into oocytes (Coffman et al. 2016). However, the timing of onset of these ring channels and the mechanisms that contribute to their maintenance through time are not well understood.

Given the challenges involved in visualizing actin dynamics in these in vivo systems, complex simulations of actin–myosin networks as proposed in Popov et al. (2016) provide useful tools for studying the dynamics and remodeling of these cellular structures. Filament contractility and alignment in these agent-based modeling simulations have been assessed by calculating the network radius of gyration and the orientational order parameter for all actin monomer units (segments) simulated (Popov et al. 2016). However, questions related to the timing and maintenance of ring channel formation as well as to their regulation by cytoskeletal proteins remain unanswered.

1.2 Topological Data Analysis for Time-Dependent Data

We investigate simulations of actin–myosin interactions where we qualitatively observe the emergence of one global hole in the simulation domain. When extracting topological information from this data, we are therefore interested in the most significant 1-dimensional hole corresponding to a ring channel (rather than other noisy features). Tracking topological features and their evolution through time is a natural question in this context; several studies have addressed aspects of this question, but not for our setting. For example, the crocker (Contour Realization of Computed k -dimensional hole Evolution in the Rips Complex) plots developed in Topaz et al. (2015) investigate time dynamics using topological data analysis. Crocker plots keep track of Betti numbers of point clouds generated by dynamical systems models of biological aggregations. The authors show that their proposed quantitative and visualization tools have predictive value in selecting models describing agent interactions (Ulmer et al. 2019). While these visual tools have been an inspiration for this work, we are interested in tracking the birth and death scale for a significant hole (ring) over consecutive time frames of the actin–myosin simulations, rather than reporting the number of topological holes with time. Our method therefore focuses on connecting features through time using topological summaries called persistence diagrams.

A much greater theoretical challenge is to assess the continuity of a topological object through time in the original image space. Several studies addressing this challenge have been proposed and adapted to data in other biological systems. The approach in Cohen-Steiner et al. (2006) introduces vineyards, which are time-parameterized stacks of persistence diagrams requiring computation of simplices (using sublevel sets filtration) at each time point. These methods are applied to protein folding trajectories in Cohen-Steiner et al. (2006). The study of Kim et al. (2020) views dynamic data sets as time-varying graphs and extracts summaries of their clustering features, with potential applications to swarming behaviors. Our work is an approxima-

tion of such theoretical approaches that is accurate for extracting the most significant feature and its emergence through time. This approach generates a path (corresponding to a vine in Cohen-Steiner et al. (2006)) using a computationally efficient algorithm based entirely on topological summaries of the data (persistence diagrams).

Since the actin–myosin simulations studied here consistently show the emergence of a global hole, our method for tracking a significant topological feature through time is complementary to this observation. This method provides an intuitive computational approach to connecting through time significant features within a persistence diagram corresponding to persistent topological holes. The resulting time-dependent path allows us to determine whether and when a structure corresponding to a ring channel emerges. Our setting applies to data represented as point clouds naturally or extracted by sampling from filamentous networks. In this work, we address how the density of sub-sampling of points from filaments affects conclusions related to the timing of ring formation in polymer networks. The proposed technique provides insights into the timing of higher-order structure formation and organization in time series data representing actin–filament interactions.

1.3 Statistical Significance in Topological Data Analysis

This work raises the natural question of when topological features become statistically significant, which is an outstanding problem in statistical topological data analysis. One of the most prominent examples of this kind of work is a series of papers by Fasy, Chazal, and others (Chazal et al. 2013; Fasy et al. 2014; Chazal et al. 2017; Bobrowski et al. 2017). In these works, the authors generally assume that there is a static compact set S on which a probability distribution P is supported. The goal is to infer the topological structure of S by analyzing a finite number of samples from P . Fasy et al. (2014) established a notion of confidence sets for persistence diagrams—the idea that the true persistence diagram will fall within a threshold distance of an observed persistence diagram, say, 95% of the time. In establishing these confidence sets, the authors observe that when the distance between the true persistence diagram and the observed one is sufficiently small, then any topological features with persistence less than the threshold value cannot be a feature of S . These features are therefore considered noise that arises from the finite-size sampling of P . Throughout this work, we will call such features “spurious,” “not significant,” or simply “noisy” depending on the context.

The notion of confidence sets requires that there is a “true” underlying persistence diagram that is to be estimated. Our challenge is slightly different in that the topological features are emergent from a dynamic filament network and the structures are fully qualitative. That is to say, there is no actual hole in the network, but often the appearance of a hole is unmistakable to the eye. Moreover, the structure of the noise underlying spurious topological features in our simulations is different from that assumed in much of the theoretical literature. In previous studies, the spurious features either arise from the discrete sampling from P or from a combination of this sampling and some model for outlier samples (with an explicit outlier probability distribution) (Chazal et al.

2017). In our case, the noise comes from filaments that are not a part of the emergent structure and their distribution is difficult to model.

There are yet more perspectives on statistical aspects topological data analysis. For example, there is work that does not use persistence diagrams as a central summary object. Notably, Bubenik (2015) proposed the notion of persistence landscapes in order to create a summary that is well-defined in a Hilbert space, leading then to a notion of p -values in that space. Blumberg et al. (2014) studied distributions of barcodes as a basis for understanding hypothesis testing and confidence sets. Recently, Maroulas et al. (2020) retained the persistence diagrams as the summary tool, but took a Bayesian perspective on inference.

In this work, the topological structure is simple, in the sense that we are looking for at most one hole; however, the analysis is difficult due to the filamentous nature of the point cloud and the unusual structure of noise that arises from the stochastic dynamics. As a result, the simplicity of structure results in visual summaries that tell a very clear story; but, the complexity of the noise interferes with a direct application of existing theory. In light of these considerations, we present three different methods for addressing the question of what constitutes *significance* of a topological feature in these simulations. In the end, these different perspectives roughly align in terms of the threshold for significance that they produce.

First, we take a hypothesis testing perspective, which requires the establishment of an appropriate null model for randomly distributed filament networks. Using computational methods, we essentially employ a filamentous version of the Poisson spatial process null hypothesis that has been studied by Bobrowski et al. (2017) and Bobrowski and Kahle (2018). In a second approach, we directly study the distribution of topological features that arise from simulations where filaments interact with molecular motors. The empirical cumulative distribution function has a plateau over an interval of persistence lengths that can be used to qualitatively infer a distinction between significant and spurious features. Finally, in a third perspective, rather than considering the simulations frame-by-frame (and pooling across frames when appropriate), we study the time-dependent paths themselves to assess significance. We do this by calculating the persistence diagram trajectory (approximating the vine concept developed in Cohen-Steiner et al. (2006)) associated with the topological feature that has maximal persistence across an entire simulation and comparing it to frame-by-frame maximally persistent features. The persistence level at the emergence time provides an estimate for significance that is similar to the first two methods.

1.4 Overview

The manuscript is organized as follows. In Sect. 2.1, we describe the MEDYAN simulation framework for actin–myosin interactions (Popov et al. 2016) and give an overview of the simulation databases generated and used in this work. In Sect. 2.2, we provide a review of the key topological data analysis concepts that we use, including the persistence diagram summary. We discuss the algorithm for tracking topological features through time in Sect. 2.3 and illustrate its application to a sample simulation. In Sect. 2.4, we describe three methods for determining a significance threshold in persis-

tence diagrams and show their agreement for time-series data from our actin–myosin simulation databases. In Sect. 3, we show how the proposed methods distinguish between and provide insights to simulations of actin–myosin interactions with different motor binding parameters. We conclude with a short discussion in Sect. 4.

2 Methods

2.1 Stochastic Simulation Framework for Actin–Myosin Interactions

2.1.1 MEDYAN Simulation Framework

For simulations of actin–myosin interactions, we use the MEDYAN model developed by the Papoian Lab and introduced in Popov et al. (2016). This agent-based modeling framework simulates actin filaments as interacting semi-flexible polymers in a solution with complex reaction and diffusion processes in three dimensions. The actin filaments interact with motor proteins such as myosins and with transient cross-linking proteins. The numerical method involves simulating a three-dimensional stochastic reaction-diffusion scheme for the active matter model using a spatially resolved Gillespie algorithm.

MEDYAN models chemical phenomena on a simulation space that is divided into compartments. Diffusion and molecular transport of various chemical species are modeled as stochastic jumps between compartments. For the purposes of our simulations, these dynamics include growth and shrinking of actin filaments, cross-linker and molecular motor binding, and active transport by molecular motors (walking). The model also uses a mechanical representation of the actin filament network where the filaments consist of multiple cylindrical segments (which we will refer to as *monomer units*) that simulate semi-flexible polymers with a given persistence length. The model includes various interaction potentials for filament deformations as they interact with other structures in the simulation domain. Additional information on details of the MEDYAN model and implementation can be found in Popov et al. (2016), Komianos and Papoian (2018). In most simulations and unless otherwise noted, we use a standard implementation of the model in Popov et al. (2016), which is parameterized for an actin–myosin network consisting of actin filaments, α -actinin cross-linking proteins, and non-muscle myosin IIa motor filaments.

2.1.2 MEDYAN Simulations with Varied Motors and Linkers

In order to develop an overview of actin–myosin interactions that reliably develop ring structures, we generated a collection of 35 MEDYAN simulations (with 200 time frames each) that have a fixed motor parameter set, but different numbers of linkers and motors; there are many motor parameters in the MEDYAN model framework, therefore we use the standard myosin-2 parameters that can be found in Popov et al. (2016). In our simulations, the motor numbers range from 0 to 10, while the linker numbers range from 0 to 3000 (these ranges include the standard values in Popov et al.

(2016)). We used this collection of simulations to study the distribution of persistence lengths discussed in Sect. 2.4.

2.1.3 MEDYAN Simulations with Varied Binding Rates

In order to assess the impact of motor binding rates on ring formation, we generated two more collections of simulations. The parameter we vary (which we refer to as the on-rate) is the rate constant of the binding reaction linking a myosin motor to two actin filaments. In the standard MEDYAN simulations in Popov et al. (2016), this on-rate has a value of 0.2 s^{-1} . We generate 40 simulation runs with an increased on-rate (0.4 s^{-1}) and 40 runs with decreased on-rate (0.1 s^{-1}). Qualitatively, we observed that the small on-rate reliably produced a ring structure in the experimental window, but the high on-rate did not. These are the collections of simulations that we study in Sect. 3 when we put our proposed methods into practice.

2.2 Topological Data Analysis

We analyze data from simulations of actin–myosin interactions using tools from topological data analysis. We give a brief introduction to the ideas behind these tools, but we refer the reader to Topaz et al. (2015) for a nontechnical overview or Ulmer et al. (2019) for a more detailed technical explanation of these techniques.

2.2.1 Persistent Homology

The data in our study consist of points in three-dimensional space that are extracted from the actin filaments in our simulations. Specifically, these data points correspond to x , y , and z locations of the cylindrical monomers that make up the actin polymers. We consider this finite set of points as a sampling from the ambient space \mathbb{R}^3 . Given this discrete set of points S , we build the Vietoris-Rips simplicial complexes. In our application, the distance between points is simply the Euclidean distance between the three-dimensional locations of the points (which are viewed as vertices or nodes). For each real-valued parameter $\epsilon > 0$ (called proximity parameter), the Vietoris-Rips construction forms a k -simplex whenever $k + 1$ points are pairwise within the distance ϵ . For example, a 1-simplex (an edge) forms when two points are within distance ϵ of each other, and a 2-simplex forms when any pair of points from a set of three points are within distance ϵ of each other. There are several simplicial complex constructions that one may choose for point clouds extracted from data (Otter et al. 2017). As in Topaz et al. (2015), we choose to use the Vietoris-Rips complex given that it is more computationally tractable than the corresponding Čech complex (Ghrist 2008). While the Čech complex provides a simplicial complex model that is faithful to the topology of the starting point cloud, the Vietoris-Rips complex approximates the Čech complex and, since it is a flag complex (maximal simplicial complex built from the underlying graph), it allows for efficient storage as a graph (Ghrist 2008).

The Vietoris-Rips complex \mathcal{S}_ϵ at scale $\epsilon > 0$ is defined as $\mathcal{S}_\epsilon = \{\sigma \subseteq S \mid d(x, y) \leq 2\epsilon \forall x, y \in \sigma\}$ (Otter et al. 2017). We are interested in using homology, a tool from

algebraic topology, to count and record features such as connected components, holes, and voids (trapped volumes) in the resulting space S_ϵ of simplicial complexes. By imposing an algebraic structure on S_ϵ , one can define the k th homology $H_k(S_\epsilon)$. The dimension of $H_k(S_\epsilon)$ is denoted as the Betti number b_k and gives the number of k -dimensional holes. For example, the number of connected components is b_0 , the number of topological holes is b_1 , and so on.

For any given a value of the proximity parameter ϵ , one can gain some insight for qualitative features of the data by building the associated simplicial complex S_ϵ and calculating its homology (Otter et al. 2017). However, any single choice for ϵ can obscure key features. The key idea underlying persistent homology is to assess topology across all reasonable values of ϵ . As ϵ increases, simplices are added to the complexes, so that there is an inclusion of complexes for smaller ϵ into those arising from a larger ϵ scale. This gives a finite sequence of nested subcomplexes which forms a filtered simplicial complex. Persistent homology then records how the homology of the simplices changes as the proximity parameter increases, thus identifying features that persist across a range of ϵ . In addition, persistent homology associates a lifetime interval to each feature, which corresponds to the range of parameters ϵ over which the feature persists. Here, we focus on the persistence diagram visualization of persistent homology computations, which allows us to keep track of these lifetime intervals by tracking their birth and death coordinates. As we further explain later, Betti numbers only provide the total number of features at each scale ϵ ; the persistence diagram visualization allows us to focus on individual features and their persistence, therefore conveying useful information for our application.

Figure 1d shows an example of a finite filtered simplicial complex that forms as we vary the value of the proximity parameter ϵ for data points (Fig. 1b) extracted from a simple MEDYAN simulation of actin polymer dynamics (Fig. 1a). The point cloud represents a sampling of cylindrical segment (monomer unit) locations along the simulated filaments; for simplicity, we extract and visualize only the first, middle, and last monomer unit location for each filament in Fig. 1. For all remaining simulations in this study, we sample 30% of the units along each filament. In “Appendix A2”, we illustrate the impact that sampling different percentages of the actin monomers in simulated filaments has on our proposed methods.

2.2.2 Persistence Diagrams and Previous Analysis of Time-Series Data

One common method introduced by Edelsbrunner et al. (2002) for displaying information about the persistent homology of a set of data points is a persistence diagram, as shown in Fig. 1c and in Fig. 2c. A persistence diagram is a multi-set of birth-death pairs corresponding to the birth and death of the homology generators. In other words, this visualization shows the scale ϵ at which a feature appears on the x -axis (birth radius) and the scale at which the feature disappears on the y -axis (death radius). We emphasize that while the terms “birth” and “death” connote time dependence, they in fact refer to the spatial scale over which a feature persists for a point cloud extracted from data at a fixed time point. Features that correspond to connected components (0-dimensional holes) always start at $\epsilon = 0$. Features that correspond to loops (1-dimensional holes) are consistently above the main diagonal. Typically the most

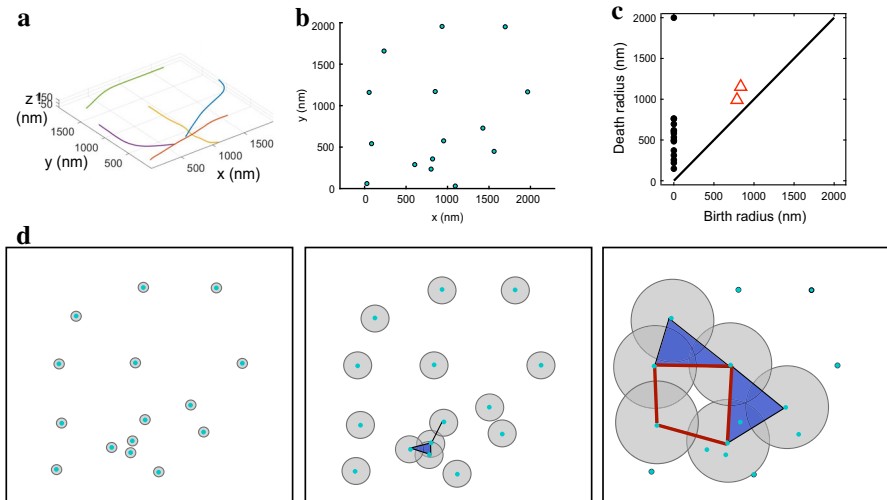


Fig. 1 Point cloud from actin simulations and resulting simplicial complexes. **(a)** Sample simulation with five actin filaments in a 2000 nm by 2000 nm by 200 nm domain. **(b)** The point cloud extracted from the simulation in **(a)**, consisting of the center and ends of each filament. The points are projected in two dimensions for visualization (and not for analysis) purposes. We later argue for more dense sampling of the filaments in obtaining the point clouds. **(c)** The persistence diagram corresponding to the point cloud in **(b)**. Black circles in the persistence diagram correspond to connected components, while red triangles correspond to 1-dimensional holes. **(d)** The resulting Vietoris-Rips complexes of the 15 points in the point cloud in **(b)** as the proximity parameter ϵ varies. Each point represents a 0-simplex, an edge is a 1-simplex if the $\epsilon/2$ circular neighborhoods of the two points intersect, a triangle is a 2-simplex if the vertices are pairwise connected by edges, etc. For ease of visualization, the right panel only shows the balls around six points in the domain and highlights in red a topological hole. Note that the complexes are visualized in two dimensions whereas the calculation in **(c)** and throughout the manuscript is carried out for the three-dimensional point clouds

significant features are those corresponding to birth-death pairs that are farther away from the diagonal, because they persist over a wider range of scales, although this is not always the case (Feng and Porter 2019). In our work, we generate persistence diagrams using the `ripsDiag` function in the TDA package in R, which calculates the Rips filtration built on top of a point cloud (Fasy et al. 2014). In particular, we use this function with the GUDHI C++ library for computing persistence diagrams (Maria et al. 2014).

In many applications, it is informative to understand the topology of the data as it varies with time. The studies of Topaz et al. (2015), Ulmer et al. (2019) propose an approach to calculate and visualize the Betti numbers of dynamic data as a function of both the proximity parameter ϵ and time. Their method, called Contour Realization of Computed k -dimensional hole Evolution in the Rips Complex (crocker), keeps track of Betti numbers $b_k(\epsilon, t)$. The authors then use this matrix of data as feature vectors that help select appropriate models of biological aggregation for given experimental data in Ulmer et al. (2019). This approach of tracking the total numbers of topological features of each dimension across proximity scale and simulation time also helps identify group phenomena such as alignment and clustering (Topaz et al. 2015).

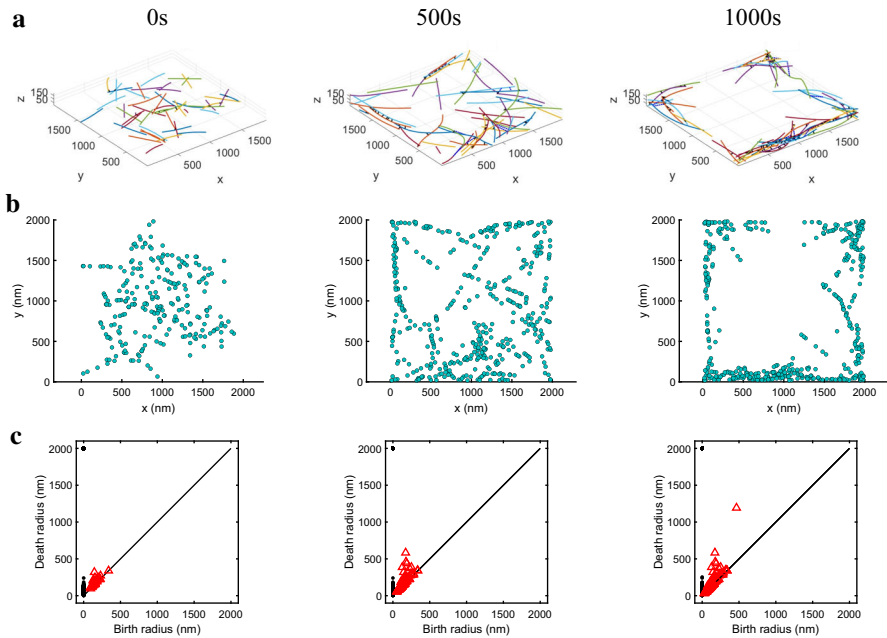


Fig. 2 Time series data and persistence diagrams. **(a)** Sample time-series data of the actin–myosin interactions in a MEDYAN simulation; different actin filaments are depicted as long colorful polymers and cross-linkers are shown as short black lines. Myosin motors are omitted in this visualization for clarity; they are represented by medium-length dashed blue lines in the supporting videos (Online Resources 1 and 2). **(b)** Sampling of 30% of the monomer units along the 50 actin filaments in each time snapshot of the simulation. **(c)** Corresponding persistence diagrams generated by calculating the Rips filtration for each of the point clouds in **(b)**; black circles correspond to connected components and red triangles correspond to loops

Our motivating scientific question is slightly different. Rather than being interested in the number of topological features, we are focused on identifying the emergence and duration of a single major feature. So, instead of counting the number of features at each time and persistence scale (as in crocker plots), we keep track of the birth and death coordinates in the persistence diagram corresponding to 1-dimensional hole features at each time point. Recording both coordinates of these points (birth-death pairs) in the persistence diagram allows us to connect through time pairs that are close in this space and to identify statistically significant features and properties of a time-changing point cloud. As discussed in the introduction, this goal is much closer to the concept of vines and vineyards developed by Cohen-Steiner et al. (2006) and more recently addressed by Kim et al. (2020).

2.3 Topological Data Analysis for Detecting Rings in Time-Series Data

Since we are interested in the formation of ring structures, we focus on the red triangles in the persistence diagrams of Fig. 2c. This figure shows three time frames in a simulation with 50 actin filaments interacting with myosin motors and α -actinin on a

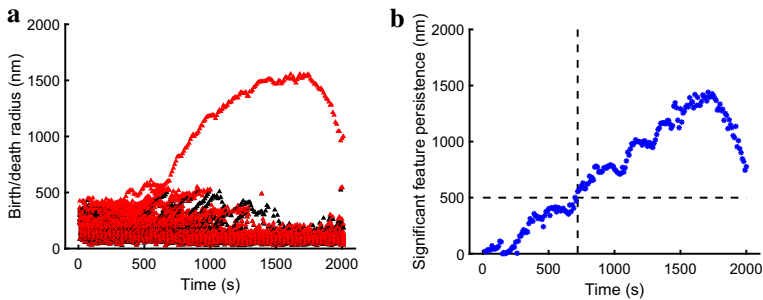


Fig. 3 Visualization of birth-death proximity parameter pairs with time in a dynamic simulation and the path generated using our method. **(a)** Black triangles denote the birth radii (ϵ_{birth}) corresponding to the formation of a topological hole, while red triangles denote the death radii (ϵ_{death}) corresponding to the disappearance of this structure. **(b)** The displayed path is the output of our method: blue dots correspond to the significant feature's persistence ($\epsilon_{\text{death}} - \epsilon_{\text{birth}}$) as a function of time. Dashed lines correspond to the statistically determined time of hole formation (vertical) and the statistically determined significant distance from the diagonal (horizontal)

domain of realistic size for actin structure organization (see Fig. 2a). For the simulations in this study, we extract 30% of the actin monomer units along each filament to generate the point cloud at each time step. In “Appendix A2”, we discuss how coarser and finer samplings affect our results. We seek a method to determine the emergence of a significant topological hole, as illustrated by the red triangle farthest from the diagonal in the third column of Fig. 2c, corresponding to 1000 s into the simulation.

2.3.1 Persistence of Rings in Time-Series Data

Persistence diagrams illustrate the birth radius ϵ_{birth} on the x -axis and the death radius ϵ_{death} on the y -axis corresponding to each topological feature, but do not allow for visualization of features across time. In order to track the evolution of these birth-death pairs in the persistence diagram over time, we constructed a visualization that overlays successive $(\epsilon_{\text{birth}}, \epsilon_{\text{death}})$ pairs for all such features as they vary in time for the dynamical systems models of actin–myosin interactions considered. Figure 3a shows an example of this visualization for the simulation in Fig. 2a, continued up to 2000 s. The black triangles correspond to radii ϵ_{birth} corresponding to the scale at which a 1-dimensional hole arises, whereas the red triangles denote radii ϵ_{death} corresponding to the scale at which the feature disappears. The hole that forms in the middle of the simulation domain is easily identified as the continuous evolution of a pair of $(\epsilon_{\text{birth}}, \epsilon_{\text{death}})$ that diverges from other features, and remains an outlier. It is worth noting that most pairs of the birth and death proximity parameter likely amount to topological noise and are close to the diagonal in Fig. 2c, or have almost overlapping birth and death radii in Fig. 3a. Figure 3b shows the output of our method (generated as described in Sect. 2.3.2): the extracted path through time for the most significant 1-dimensional feature in Fig. 3a. The displayed path corresponds to the difference between the specific pairs of death radii (red triangles in Fig. 3a) and birth radii (black triangles in Fig. 3a) for the path identified using our proposed algorithm.

2.3.2 Algorithm for Visualization of Ring Structure Persistence

To explore the emergence of these continuous ring structures in time-series data, we propose a method for connecting the pairs of $(\epsilon_{\text{birth}}, \epsilon_{\text{death}})$ that are most likely to correspond to the same significant 1-dimensional hole structure through time as illustrated by the pair emerging around time 600 s in Fig. 3a. In the persistence diagram plots in Fig. 2c, our method will connect the red triangles corresponding to consecutive time points (at intervals of 10 s). Our approach consists of the following steps:

1. Calculate all birth-death pairs $(\epsilon_{\text{birth}}, \epsilon_{\text{death}})$ corresponding to the dimension of interest for each time step.
2. At each time step, order the birth-death pairs from highest persistence to lowest persistence (i.e., order them in decreasing order of $\epsilon_{\text{death}} - \epsilon_{\text{birth}}$).
3. Start by considering the first two time steps and calculate the matrix of L_∞ distances between all birth-death pairs at the first time step and the pairs at the next time step:

$$\begin{aligned} & \|(\epsilon_{\text{birth}1}, \epsilon_{\text{death}1}) - (\epsilon_{\text{birth}2}, \epsilon_{\text{death}2})\|_\infty \\ &= \max(|\epsilon_{\text{birth}2} - \epsilon_{\text{birth}1}|, |\epsilon_{\text{death}2} - \epsilon_{\text{death}1}|). \end{aligned} \quad (1)$$

4. For each pair in the first time step and starting with the most persistent feature, find the pair in the second time step that is closest to it in the sense of the L_∞ metric in (1). If this smallest distance is greater than a parameter that we denote as the linkage tolerance d_ℓ , then the pair is not counted as a connection.
5. Repeat steps 2–4 for all consecutive time steps until the end of the simulation. Pairs at consecutive time points are then combined into paths (see Fig. 7 in “Appendix A1”).
6. Isolate the most significant path by finding the pair that is farthest from the diagonal ($\epsilon_{\text{birth}} = \epsilon_{\text{death}}$) at some time point during the simulation and extracting the time path corresponding to this pair.

The L_∞ distance between points in persistence diagrams corresponding to consecutive time frames (Step 3) is the same distance used to calculate the bottleneck distance between two persistence diagrams (however, using the L_2 distance metric does not qualitatively change the output of our algorithm). The bottleneck distance is the L_∞ cost of the optimal matching between points in two persistence diagrams. In practice, computational approximations are used to perform this matching. In our setting, we prioritize pairing the most significant (persistent) features through time. Given the small time step in our simulations (10 s), we expect small changes in the actin monomer unit locations between consecutive time frames, and we use the observation that such small changes in point clouds also result in small changes in the corresponding persistence diagrams (by the stability theorem in Cohen-Steiner et al. (2007)). By ordering the features from high to low persistence in Step 2 and by finding the nearest neighbor of the most persistent features in the sense of the L_∞ metric in Steps 3–4, we take advantage of the intuition behind the stability theorem and expect our pairing for the most significant feature to be accurate. The linkage tolerance d_ℓ for our simulations is chosen to be 370 nm based on the scale of the spatial domain of interest. Since in

Step 1 of the algorithm we calculate all birth-death pairs, one option is to select the pair with the largest persistence and only calculate its corresponding path as outlined in step 6, followed by steps 2–5. As presented, the algorithm outlines the calculation of all paths through time in the persistence diagram. Sample code is provided in the repository on GitHub (2020).

The connections between points (birth-death pairs) that are close to the diagonal in the persistence diagram representation (see Fig. 7) may not necessarily correspond to the same 1-dimensional topological hole generated from the data at later times. On the other hand, the algorithm accurately matches through time the birth-death pairs that are farthest from the diagonal in the persistence diagram, since a single persistent path corresponding to a 1-dimensional hole emerges (see Figs. 3 and 7b).

It is worth noting that the standard parameters driving the dynamics of the actin–myosin simulations considered here lead fairly consistently to at most one global emerging hole (ring) in the simulation domain, therefore the significant path extracted in step 5 of the algorithm likely corresponds to the same persistent 1-dimensional hole in the data. Given two birth-death pairs that cross or come close together in persistence diagrams for consecutive simulation frames, our algorithm could lead to uncertainty in whether the paths extracted are properly separated. In addition, if two birth-death pairs extracted at one time point of the simulation are closest to only one pair in the next time point, the algorithm could choose an incorrect match for a path. However, given the dynamics of the actin–myosin network in our simulations and the global 1-dimensional hole structures we seek, this is unlikely to occur in our application. The exception to this is the matching of points close to the diagonal in the persistence diagrams, which are nonetheless below the significance threshold we discuss in the next section.

In “Appendix A1”, we illustrate in Fig. 7 the results of connecting the birth-death pairs corresponding to 1-dimensional topological hole features for consecutive times in the simulation of Figs. 2a and 3a. Each of the rainbow-colored paths in the persistence diagram in Fig. 7a connects through simulation time the birth-death pairs corresponding to a 1-dimensional hole feature using our algorithm, while Fig. 7b extracts the significant path corresponding to the largest hole that emerges in the simulation domain (see also the animation in Online Resource 1). The proposed method also generalizes to connecting birth-death pairs within a persistence diagram for higher-order cavities. The video in Online Resource 2 illustrates both the significant path for the 1-dimensional hole (dark orange) and all paths corresponding to the 2-dimensional voids (blues and greens) forming through time in a standard simulation where we extract three monomer unit locations per filament, as shown in Fig. 1b. In this example, all cavities (2-dimensional holes) have small persistences while a 1-dimensional hole clearly emerges, indicating that a thin tunnel emerges from the actin–myosin network in the simulation domain.

Motivated by investigating the timing of ring structure formation, we consider an alternative visualization to Fig. 7b that plots the persistence $\epsilon_{\text{death}} - \epsilon_{\text{birth}}$, i.e., the vertical distance from each birth-death pair in the significant path to the diagonal in the persistence diagram. This visualization shows rotated persistence diagrams that are visualized in a birth-persistence coordinate system (Adams et al. 2017; Stolz et al. 2018) with $(\epsilon_{\text{birth}}, \epsilon_{\text{death}} - \epsilon_{\text{birth}})$ coordinates. Figure 3b shows the evolution of the

feature persistence as a function of time in a sample simulation. While this is not the case for all applications (Feng and Porter 2019), we will consider pairs that are farther away from the diagonal in persistence diagrams, and thus higher along the vertical axis in Fig. 3b, to correspond to 1-dimensional hole features that are more significant. We proceed to discussing methods for assessing feature significance in this context in Sect. 2.4.

2.4 Estimating Significance for Topological Features

This work raises the question of how we might distinguish significant features from noise in time-series data. A common method for identifying signal in persistent homology is to seek features that have a large persistence (i.e., $\epsilon_{\text{death}} - \epsilon_{\text{birth}}$) and therefore correspond to birth-death coordinates farther away vertically from the diagonal. This is, of course, not universally the right thing to do. Recent studies with applications to neuroscience and voting maps have addressed the limitations of this intuition since some short-lived features may hold key insights and the interpretation of feature persistence is not always clear (Stolz et al. 2017; Feng and Porter 2019). In our application, large persistence is a meaningful notion. The point clouds we consider are sub-sampled from the simulated three-dimensional locations of actin monomer units, and Vietoris-Rips simplicial complexes are constructed based on the Euclidean pairwise distances between these points. It follows that persistent 1-dimensional features (those that are farther from the diagonal in the persistence diagram) correspond to large-scale holes emerging in the simulation domain, for which it is our challenge to identify whether and when they form and how long they last (see the video in Online Resource 1).

We now describe three methods by which we selected the threshold for significance of feature persistence in our application. The first two methods are statistical, while the third is trajectory-based and relies on the observation that there is a single emerging hole in the domain of our simulations. The methods outlined below give similar estimates for the threshold level for feature persistence, which we conservatively choose to be 500 nm. Given this threshold, we define the onset of ring formation as the last time that the main feature's persistence crosses from below to above the significance threshold, as illustrated by the vertical and horizontal dashed lines in Fig. 3b.

2.4.1 Significance Through Hypothesis Testing

In the hypothesis testing perspective, we need to establish an appropriate null model for randomly distributed actin filaments. By analyzing the distribution of the maximum persistence lengths in the null model frames, we will establish a threshold so that the null model will be unlikely to generate features with persistence higher than this threshold.

To develop the null model, we first record the filament lengths and the number of monomer units extracted from each filament at each time frame from the MEDYAN simulation database described in Sect. 2.1.2. We then generate a null frame by randomly choosing a MEDYAN-simulated frame and by constructing straight filaments with the same lengths and the same density of sampling of the actin monomer units as

in the original frame. These filaments are assigned random positions (drawn uniformly at random from the simulated region) and orientations (with angle off the x -axis drawn uniformly from the interval $[0, 2\pi]$); we reject filaments that extend outside the simulation domain (see Fig. 9a in “Appendix A3” for an example null frame). In this way, we generate 1000 null frames that have the same filament numbers, lengths, and the same numbers of monomer units sampled from the polymers as in model simulation frames generated in Sect. 2.1.2. We compute the persistent homology for the point clouds extracted from each null frame; see Fig. 9b in “Appendix A3” for an example of a persistence diagram corresponding to a null frame. In Fig. 4a, we study the distribution of the maximum persistence from each of the 1000 null frames and find that the .99 quantile of this distribution is roughly 312 nm; this means that there is less than 1% chance that a random Poisson spatial process would generate a 1-dimensional hole of this (or larger) size. Therefore, a significance threshold larger than or equal to this level is suitable for rejecting the null hypothesis that the feature can be generated by this type of randomly drawn filament network. As we see below, this method produces a small estimate for the significance threshold, most likely because this model for noise is not rich enough to capture everything that produces spurious features in the data set. This is, in part, why we use multiple perspectives on assessing significance in this work.

2.4.2 Significance Through Analysis of Spurious Feature Distribution

In the second statistical method, we pool all the persistence lengths from all time frames in the database of MEDYAN simulations described in Sect. 2.1.2 and compute the corresponding “survival function” (the probability that features have a persistence larger than a given fixed value). Figure 4b illustrates this survival function, where the y -axis values are shown in logarithmic scale and represent the proportion of features with persistence size larger than the corresponding persistence length on the x -axis. We notice the emergence of a plateau which corresponds to the transition from spurious to significant topological features of interest (marked with stars in Fig. 4b). In these results, we use a sampling density of 30% of the actin monomers; we further comment on the impact of the sampling density in “Appendix A3”.

We similarly compute the survival function (marked with x 's in Fig. 4b) corresponding to the null model frames generated as described in Sect. 2.4.1 (see Fig. 9a for a sample null model frame and Fig. 9b for its corresponding persistence diagram). The survival functions for the frames generated by the null model and the MEDYAN simulations have considerably different shapes. The emergent plateau that begins somewhere between 300 and 500 nm in the MEDYAN-generated frames marks a transition from spurious topological features to values that are associated with the topological features of interest. This figure shows that there are very few 1-dimensional features with a persistence level in this interval that are generated by the random filament network in the null model (x 's in Fig. 4b). This is consistent with our findings in Sect. 2.4.1 based on the null model alone.

At the suggestion of an anonymous reviewer, we also include the distribution of topological features from MEDYAN simulations if we remove the maximally persistent feature in each persistence diagram. This removal of points ensures that we are

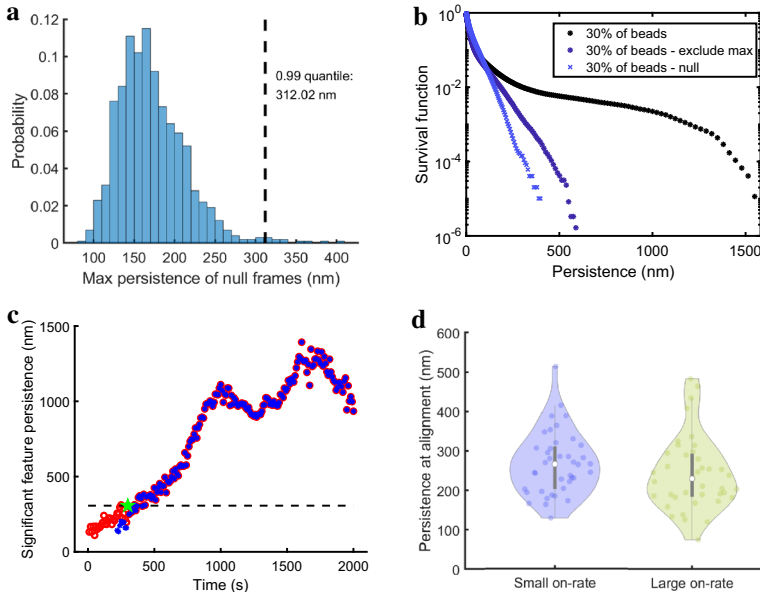


Fig. 4 Assessing significance of topological features. **(a)** Distribution of the maximum persistence in each generated null frame. Dashed vertical line indicates the .99 quantile for this distribution. **(b)** Survival function, i.e., proportion of features with persistence size larger than the corresponding persistence length on the x -axis, for the database of 35 MEDYAN simulations (black stars), for the same simulations but excluding the maximum persistence at each time frame (purple stars), and for the null model frames (blue x 's). The y -axis is on a log scale for ease in visualization. **(c)** In red, we mark the persistence level corresponding to the 1-dimensional topological feature with maximum persistence at that time. In blue, we mark the persistence corresponding to the path of the most significant hole through time. There is a time point at which the two paths align (marked by a green star). We denote the persistence level at that time point as the persistence at alignment and mark this level using the horizontal dashed line. **(d)** Violin plot of the persistence level at alignment (see panel c) for 40 small on-rate simulations (blue) and 40 large on-rate simulations (green)

not including topological features that have emerged from the spurious feature cloud, while also including noisy effects missed in the null model. The associated survival function, marked with purple stars, shows a clear departure from the other two, but is consistent with our selection of 500 nm as a significance threshold.

2.4.3 Significance Through Path Tracing

In the third approach, we propose a means of finding the persistence length at which the most significant path (as found using the algorithm in Sect. 2.3.2) emerges as the dominant feature in a video of actin–myosin interactions. For each model simulation in the collection described in Sect. 2.1.2, we compare the maximum persistence among all 1-dimensional features at each time to the time-dependent persistence of the significant path. In Fig. 4c (corresponding to a sample simulation with standard parameters), we mark in red the maximum persistence level among all 1-dimensional topological features at each time. We also mark in blue the persistence corresponding to the path of the most significant hole through time.

Figure 4c shows that, after a period with noisy, short-lived loop features, these two time-dependent persistence paths align, thus providing an indicator of the threshold for significance. In Fig. 4d, we visualize the persistence levels at the alignment of these paths (see green star and horizontal dashed line in Fig. 4c) using violin plots for the simulation sets of small and large myosin motor binding rates described in Sect. 2.1.3 (and further analyzed in Sect. 3). The mean values for these persistence levels at alignment (267.7 nm for small on-rate and 241.9 nm for large on-rate) provide estimates for the significance threshold and have qualitative agreement with the threshold values predicted using the previous methods.

3 Analysis of a Filamentous Network Model

To illustrate the power of the methods introduced in Sect. 2 in distinguishing between different dynamic behaviors, we consider the simulation sets described in Sect. 2.1.3, where we focus on actin–myosin organization in the context of large and small on-rates (motor binding rate parameter).

Figure 5 shows snapshots of the final actin configurations in two sample stochastic runs of simulations with these parameters, as well as the emergence of ring structure with time (as introduced in Fig. 3b) in each case. A decreased on-rate leads to alignment of the actin filaments at the boundaries, and thus a clear hole emerges in the simulation domain (Fig. 5a). A large on-rate leads to more frequent interactions between actin filaments and myosin motors and as result more contractile behavior and clustering of filaments in various regions of the domain (Fig. 5b).

This distinction in the dynamics is clear when visualizing the persistence ($\epsilon_{\text{death}} - \epsilon_{\text{birth}}$) plots in Fig. 5. The small on-rate setting corresponds to a clear hole in the simulation domain that maintains high feature persistence value (i.e., it is far from the diagonal in the persistence diagram) as time progresses in each of the two simulations illustrated. In the large on-rate case, a hole forms but is not maintained over time and therefore the significant feature does not persist throughout the simulation. When the myosin motors have a higher likelihood of binding to actin filaments, the dynamics of the polymer network shows more variability, as illustrated by outcomes from application of our technique to two stochastic realizations of the dynamics with this parameter choice in Fig. 5b. Here, the significant path captures several short-lived holes that are not maintained throughout time.

Our algorithm allows us not only to observe these distinguishing features, but also to quantify them in the 40 MEDYAN simulations of the actin–myosin interactions for each setting: half on-rate and double on-rate (in reference to the standard myosin-2 parameters in Popov et al. (2016)). For each parameter choice, we are interested in exploring patterns related to the number of emergent holes, whether they close or remain open, the lifespan of the significant holes, and their size—quantified as the maximum persistence of the topological feature during the time it is significant. As outlined in Sect. 2.4, we use a conservative significance level threshold of 500 nm for this analysis, as validated by the three methods for establishing significance. When carrying out the same analysis for a threshold of 300 nm, the qualitative conclusions

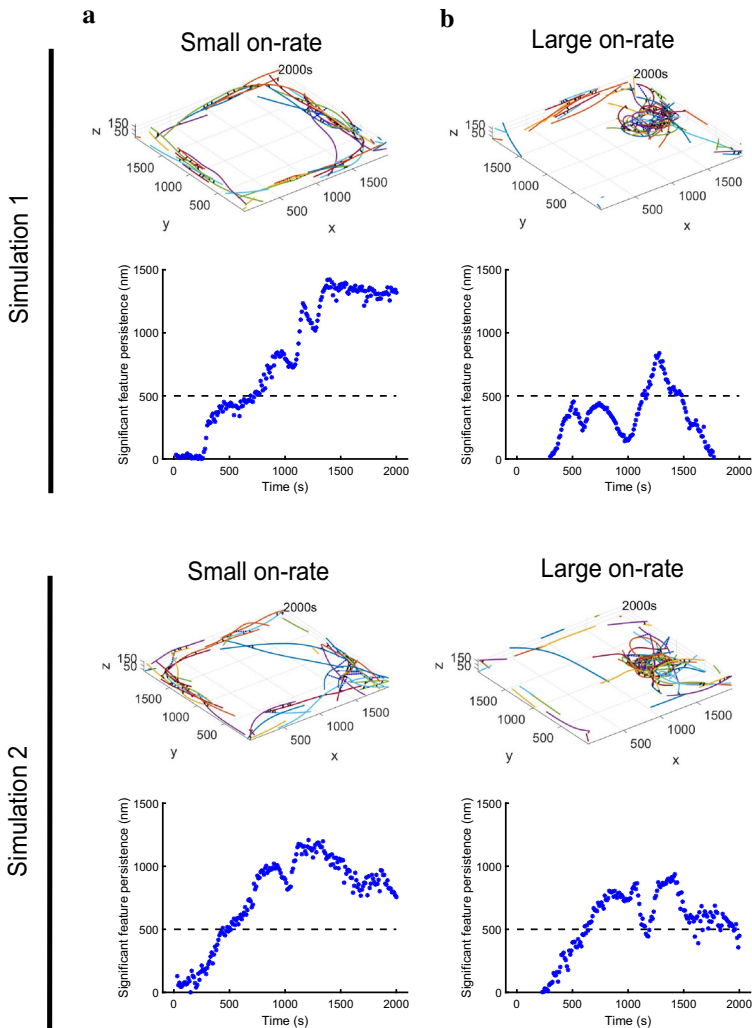


Fig. 5 Analysis of simulations with small (**a**), respectively large (**b**) motor binding rate for two stochastic simulations. Within each simulation: (Top) Distribution of actin filaments, myosin motors, and cross-linkers at the final simulation time; (Bottom) Visualization of ring emergence as time-dependent persistence of the significant path corresponding to a 1-dimensional hole

were the same: the small on-rate simulations produced larger holes more reliably than the large on-rate simulations.

In the small on-rate simulations, the dynamics is consistently characterized by alignment of filaments at the domain boundaries (see Fig. 5a). In all runs, our method identifies a significant hole that persists until the end of the simulation time, allowing us to quantify the ensemble statistics of the timing of ring formation. The estimate for the mean time of ring formation is 693 ± 61.81 s (this and all the following intervals are at the 95% confidence level). We illustrate the distribution of the timing of significant

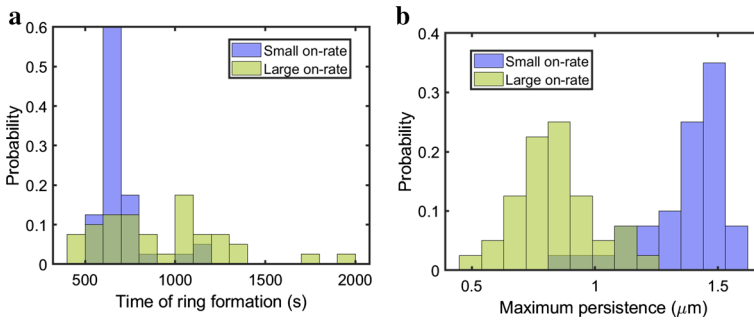


Fig. 6 Histograms of the time and persistence (size) of the first significant ring emergence in small vs. large on-rate simulations. **(a)** Histograms of the time of the first significant ring onset in blue (small on-rate) and in green (large on-rate) based on 40 simulations each. **(b)** Histograms of the ring size (maximum persistence) in blue (small on-rate) and in green (large on-rate) based on 40 simulations each

ring formation in the histogram in Fig. 6a in blue. We can describe the size of the detected rings by the maximum persistence ($\epsilon_{\text{death}} - \epsilon_{\text{birth}}$ in the persistence diagram) over the time that the ring is above the significance level. Figure 6b illustrates the histogram of the ring sizes in blue, where the estimate for the mean of the maximum persistence size achieved by significant holes in these 40 simulations is $1.36 \pm 0.07 \mu\text{m}$.

As illustrated in Fig. 5b, the large on-rate simulations lead to more contractility of the actin–myosin network and to a higher likelihood of polymer clusters forming in the domain. The estimate for the mean number of significant holes emerging throughout these simulations is 4.02 ± 1.22 (see Simulation 2 in Fig. 5b for an example of a run that yields several short-lived significant holes). The holes in the actin–myosin network identified by our method for these simulations are on average smaller and shorter-lived than those in the small on-rate simulations, with an estimate for the mean hole lifespan of 226.21 ± 88.82 s and for the mean of the maximum persistence of $0.85 \pm 0.07 \mu\text{m}$. Figure 6a shows the histogram of the distribution for the time of ring formation for the first hole emerging in the large on-rate simulations in green. In some stochastic runs, the onset time is similar to that of the significant holes in the small on-rate simulations, however the holes in the large on-rate simulations fall apart as their persistence soon goes below the significance threshold (see Fig. 5b for two representative examples). In other stochastic runs, a significant ring does not emerge until much later in the simulation, so that the variance of the time of significant hole formation is considerably larger for the higher binding rate setting. In Fig. 6b, we illustrate the histogram of the maximum persistence over all rings in each large on-rate simulation in green. As expected given that the rings are less prominent in the large on-rate simulations, this distribution is shifted to the left (corresponding to smaller maximum ring persistence) compared to the small on-rate simulations.

Therefore, our proposed approach for analyzing time-series data of cellular interactions is able to rigorously distinguish between and quantify emerging features of parameter changes in agent-based model simulations of cellular polymer interactions.

4 Conclusions

Understanding complicated interactions of filamentous networks and multiple chemical species at the cellular level often requires complex simulations that provide insight into the temporal and spatial dynamics of the interacting proteins. Here, we carried out numerical simulations of actin–myosin and crosslinker interactions using the MEDYAN model (Popov et al. 2016). Filament contractility and alignment in these models have been studied using classical tools such as calculation of the network radius of gyration and of an orientational order parameter of the system (Popov et al. 2016). However, an understanding of how filamentous networks interact to create higher-order structure and organization in cells is lacking. We propose a computational technique based on topological data analysis to identify ring structure in complex simulations of filament organization.

The method we propose requires that we sample the spatial distribution of filaments at each time point of a dynamical simulation and thus extract a point cloud. Computing the persistent homology of this discrete set of data points generates a persistence diagram, which is commonly used to represent and visualize birth-death pairs corresponding to topological objects such as loops. We present an algorithm that connects significant pairs across multiple persistence diagrams over time. As previously mentioned, Cohen-Steiner et al. (2006) introduced this concept, called vines and vineyards, defined as continuous families of persistence diagrams for time series of continuous functions. Computation of these vineyards requires a list of simplices at each time point (and relies on sublevel set filtrations), while our algorithm only requires knowledge of the persistence diagrams (birth-death pairs) at each time. Taking advantage of the relative simplicity of the dominant topological feature we pursue, our proposed approach has the advantage that it only requires topological summaries of the data. We also present multiple perspectives on significance of features in persistence diagrams. We emphasize transparent computational methods for detecting and quantifying the most significant higher-order structure that emerges from polymer network interactions and evolves in time according to a stochastic dynamical system. Individually, these perspectives provide an incomplete view of the persistence distribution of spurious topological features. Taken together, however, the resonance among these perspectives provides a robust assessment of when the dominant feature of the dynamics emerges and how long it endures.

Acknowledgements MVC was supported by The Ohio State University President's Postdoctoral Scholars Program and by the Mathematical Biosciences Institute at The Ohio State University through NSF DMS-1440386. ATD is supported by NSF DMS-1554896. RJ and SAM are supported by NIH R01GM122082-01 and by the NSF-Simons Southeast Center for Mathematics and Biology (SCMB) through the grants National Science Foundation MS1764406 and Simons Foundation/SFARI 594594. Part of this research was conducted using computational resources and services at the Ohio Supercomputer Center through the Mathematical Biosciences Institute at The Ohio State University. We also thank the Mathematical Biology Institute at The Ohio State University for supporting Riley Juenemann in attending their Summer Undergraduate Capstone Conference and thus enabling our collaboration. We are grateful to Dr. Chad Topaz, Dr. Henry Adams, Dr. Sarah Day, Dr. Peter Bubenik, Dr. Brittany Fasy and her group, and Dr. Garegin Papoian and his lab for helpful conversations in developing this work. We also thank the anonymous referee who suggested carrying out a rigorous hypothesis test for establishing feature significance.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

A1: Visualization of Birth-Death Pairs Connected as Paths Through Time in the Persistence Diagram

One approach to illustrating the results of our algorithm on a sample MEDYAN simulation is shown in Fig. 7, where we connect the birth-death pairs corresponding to 1-dimensional topological hole features at consecutive times in the simulation. The points $(\epsilon_{\text{birth}}, \epsilon_{\text{death}})$ (marked by red triangles in Fig. 2c) are connected with straight line segments in Fig. 7. The algorithm described in Sect. 2.3.2 ensures that these segments (which connect the features through time) are relatively short for each path. In Fig. 7b, we isolate the significant path corresponding to the largest hole that emerges in the simulation domain. This is achieved by identifying the path that has the largest $\epsilon_{\text{death}} - \epsilon_{\text{birth}}$ persistence at some time point in the simulation. We visualize this significant path separately in Fig. 7b. The formation of this path and its departure away from the diagonal (i.e., the addition of birth-death pairs at subsequent times) are also illustrated in the animation in Online Resource 1.

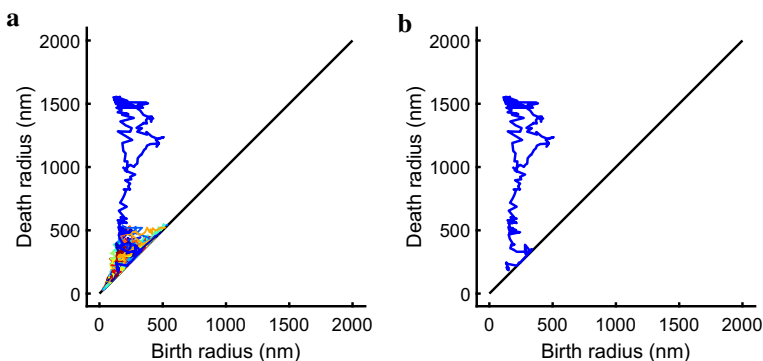


Fig. 7 Visualization of pairs of birth-death proximity as paths connected through time in the persistence diagram. **(a)** Each rainbow-colored path corresponds to a topological hole identified and connected through time using our algorithm. **(b)** The most significant path is isolated. Note that these paths are only meaningful outside a noisy threshold above the diagonal that we discuss in Sect. 2.4 on estimating a significance threshold

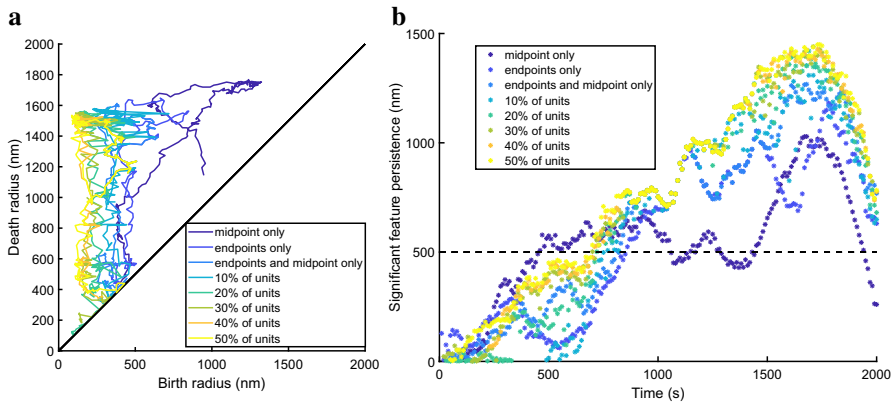


Fig. 8 Sensitivity to the sampling density of the actin filaments. **(a)** Visualization of the most significant topological hole path within a persistence diagram for an increasing fraction of actin monomer units sampled along each filament in a standard binding rate MEDYAN simulation. **(b)** Visualization of the time-dependent persistence for an increasing fraction of actin units sampled along each filament

A2: Effect of Monomer Unit Sampling on Significant Paths

We explore how the significant paths corresponding to 1-dimensional features (as in Fig. 7) and the time-dependent persistence plots (as in Fig. 3b) change with an increasing fraction of monomer units extracted from each simulated filament. Figure 8a shows that sampling more units (points) along the filaments leads to the significant paths moving left in the persistence diagram but converging at a mid-range of the sampling density. We expect this to be the case since the more units we extract, the closer the points are in the point cloud, and therefore we observe the 1-dimensional holes at a smaller birth radius. As the percentage of monomer units extracted along each filament increases, we also observe that the paths are more persistent over time in Fig. 8b. The onset of the significant hole and its maximum persistence are similar for mid-range to high sampling densities. On the other hand, only extracting the midpoint of the actin filaments leads to a significant path that is not above the persistence threshold consistently through time. As expected for persistent homology calculations, considering large point clouds extracted from a larger fraction of the monomer units along each filament increases the computational time of the algorithm proposed.

A3: Effect of Monomer Unit Sampling on Significance Through Spurious Feature Distribution

Figure 9a shows an example of a null frame generated using the framework described in Sect. 2.4.1, where actin filaments are assigned random locations and orientations in a typical simulation domain. In Fig. 9b, we illustrate the corresponding persistence diagram for the point cloud of monomer units extracted from the frame in Fig. 9a. Here, we focus on the impact of the sampling density of actin monomer units on our method for establishing significance as described in Sect. 2.4.2. We focus on two sampling

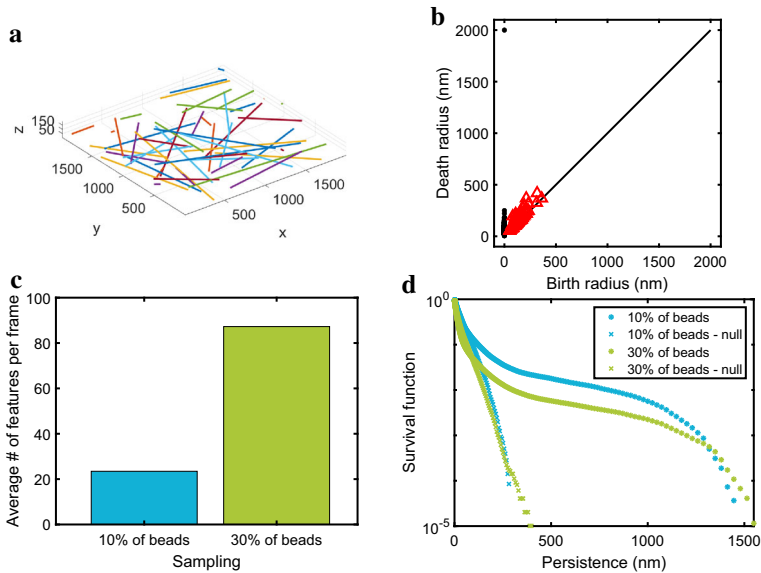


Fig. 9 Assessing significance through the distribution of persistence and comparison with the null model. (a) Sample frame of actin filaments generated according to the null model described in the main text. (b) The corresponding persistence diagram for the null model frame in (a); black circles correspond to connected components and red triangles correspond to loops. (c) Average number of 1-dimensional features per time frame for two sampling densities of the actin monomer units in our database of 35 MEDYAN simulations. (d) Survival function, i.e., proportion of features that are larger than the corresponding persistence length on the x -axis for the database of 35 MEDYAN simulations (stars) and for the null model frames (x 's). The y -axis is on a log scale for ease in visualization

densities used to generate Fig. 8, which produced similar results for the time-dependent paths (10% and 30%). In Fig. 9c, we compute the average number of topological features in each video frame. The 30% sample density produced approximately four times the number of features produced using the 10% sampling. While we expect to observe more 1-dimensional features in the larger point clouds corresponding to the 30% sampling density, we also lack a general understanding of how many noise-induced features will appear for a given simulation or for different sampling densities.

In Fig. 9d, we study the distribution of spurious topological features for both null model frames and MEDYAN model simulation frames for the two sampling densities. We recall from Sect. 2.4.2 that, to generate Fig. 9d, we pool all persistence lengths for each sampling density and compute the “survival functions” corresponding to the null model frames (x 's, see Fig. 9a for a sample null frame) and corresponding to the model-generated frames (stars, see Fig. 2a for sample simulation frames). As observed in the text, and for both sampling densities, the survival functions for the null and model-generated frames have significantly different shapes.

When comparing the survival function plots for the model-generated frames using the two sampling densities (marked with stars in Fig. 9d), we note that while the tail behavior and the proportion values differ, the emergent plateau that begins between

300 and 500 nm is present in both curves. This shows that the choice of 500 nm for the significance threshold is not dependent on the sampling density of the actin filaments.

References

- Adams H, Emerson T, Kirby M, Neville R, Peterson C, Shipman P, Chepushtanova S, Hanson E, Motta F, Ziegelmeier L (2017) Persistence images: a stable vector representation of persistent homology. *J Mach Learn Res* 18(1):218–252
- Blumberg AJ, Gal I, Mandell MA, Pancia M (2014) Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces. *Found Comput Math* 14(4):745–789
- Bobrowski O, Kahle M (2018) Topology of random geometric complexes: a survey. *J Appl Comput Topol* 1(3–4):331–364
- Bobrowski O, Kahle M, Skraba P et al (2017) Maximally persistent cycles in random geometric complexes. *Ann Appl Probab* 27(4):2032–2060
- Bobrowski O, Mukherjee S, Taylor JE et al (2017) Topological consistency via kernel estimation. *Bernoulli* 23(1):288–328
- Bubenik P (2015) Statistical topological data analysis using persistence landscapes. *J Mach Learn Res* 16(1):77–102
- Chazal F, Fasy B, Lecci F, Michel B, Rinaldo A, Rinaldo A, Wasserman L (2017) Robust topological inference: Distance to a measure and kernel distance. *J Mach Learn Res* 18(1):5845–5884
- Chazal F, Fasy BT, Lecci F, Rinaldo A, Singh A, Wasserman L (2013) On the bootstrap for persistence diagrams and landscapes. arXiv preprint [arXiv:1311.0376](https://arxiv.org/abs/1311.0376)
- Coffman VC, Kachur TM, Pilgrim DB, Dawes AT (2016) Antagonistic behaviors of NMY-1 and NMY-2 maintain ring channels in the *C. elegans* gonad. *Biophys J* 111(10):2202–2213
- Cohen-Steiner D, Edelsbrunner H, Harer J (2007) Stability of persistence diagrams. *Discrete Comput Geom* 37(1):103–120
- Cohen-Steiner D, Edelsbrunner H, Morozov D (2006) Vines and vineyards by updating persistence in linear time. In: Proceedings of the twenty-second annual symposium on Computational geometry, pp 119–126. ACM
- Dirafzoon A, Bozkurt A, Lobaton E (2016) Geometric learning and topological inference with biobotic networks. *IEEE Trans Signal Inf Process Netw* 3(1):200–215
- Edelsbrunner H, Harer J (2008) Persistent homology—a survey. *Contemp Math* 453:257–282
- Edelsbrunner H, Harer J (2010) Computational topology: an introduction. American Mathematical Society
- Edelsbrunner H, Letscher D, Zomorodian A (2002) Topological persistence and simplification. *Discrete Comput Geom* 28(4):511–533
- Fasy BT, Kim J, Lecci F, Maria C (2014) Introduction to the R package TDA. arXiv preprint [arXiv:1411.1830](https://arxiv.org/abs/1411.1830)
- Fasy BT, Lecci F, Rinaldo A, Wasserman L, Balakrishnan S, Singh A et al (2014) Confidence sets for persistence diagrams. *Ann Stat* 42(6):2301–2339
- Feng M, Porter MA (2019) Persistent homology of geospatial data: a case study with voting. arXiv preprint [arXiv:1902.05911](https://arxiv.org/abs/1902.05911)
- Ghrist R (2008) Barcodes: the persistent topology of data. *Bull Am Math Soc* 45(1):61–75
- GitHub (2020) Sample R code for connecting and visualizing paths of birth-death pairs through time in persistence diagrams generated from time-series point cloud data. https://github.com/veroniq04/Connect_birth_death_TDA_R
- Hudson AM, Mannix KM, Cooley L (2015) Actin cytoskeletal organization in *Drosophila* germline ring canals depends on Kelch function in a Cullin-RING E3 ligase. *Genetics* 201(3):1117–1131
- Kelley CA, Cram EJ (2019) Regulation of actin dynamics in the *C. elegans* somatic gonad. *J Dev Biol* 7(1):6
- Kim W, Mémoli F, Smith Z (2020) Analysis of dynamic graphs and dynamic metric spaces via zigzag persistence. In: Topological data analysis, pp 371–389. Springer (2020)
- Komianos JE, Papoian GA (2018) Stochastic ratcheting on a funneled energy landscape is necessary for highly efficient contractility of actomyosin force dipoles. *Phys Rev X* 8(2):021006
- Maria C, Boissonnat JD, Glisse M, Yvinec M (2014) The gudhi library: simplicial complexes and persistent homology. In: International congress on mathematical software, pp 167–174. Springer

- Maroulas V, Nasrin F, Oballe C (2020) A bayesian framework for persistent homology. *SIAM J Math Data Sci* 2(1):48–74
- Ong S, Foote C, Tan C (2010) Mutations of DMYPT cause over constriction of contractile rings and ring canals during *Drosophila* germline cyst formation. *Dev Biol* 346(2):161–169
- Osorio DS, Chan FY, Saramago J, Leite J, Silva AM, Sobral AF, Gassmann R, Carvalho AX (2018) Flow-independent accumulation of motor-competent non-muscle myosin II in the contractile ring is essential for cytokinesis. *bioRxiv* p 333286
- Otter N, Porter MA, Tillmann U, Grindrod P, Harrington HA (2017) A roadmap for the computation of persistent homology. *EPJ Data Sci* 6(1):17
- Popov K, Komianos J, Papoian GA (2016) MEDYAN: mechanochemical simulations of contraction and polarity alignment in actomyosin networks. *PLoS Comput Biol* 12(4):e1004877
- Robinson DN, Cant K, Cooley L (1994) Morphogenesis of *Drosophila* ovarian ring canals. *Development* 120(7):2015–2025
- Robinson DN, Cooley L (1996) Stable intercellular bridges in development: the cytoskeleton lining the tunnel. *Trends Cell Biol* 6(12):474–479
- Saggar M, Sporns O, Gonzalez-Castillo J, Bandettini PA, Carlsson G, Glover G, Reiss AL (2018) Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nat Commun* 9(1):1399
- Schwayer C, Sikora M, Slováková J, Kardos R, Heisenberg CP (2016) Actin rings of power. *Dev Cell* 37(6):493–506
- Stolz BJ, Emerson T, Nahkuri S, Porter MA, Harrington HA (2018) Topological data analysis of task-based fMRI data from experiments on schizophrenia. *arXiv preprint arXiv:1809.08504*
- Stolz BJ, Harrington HA, Porter MA (2017) Persistent homology of time-dependent functional networks constructed from coupled time series. *Chaos Interdiscip J Nonlinear Sci* 27(4):047410
- Topaz CM, Ziegelmeier L, Halverson T (2015) Topological data analysis of biological aggregation models. *PLoS One* 10(5):e0126383
- Ulmer M, Ziegelmeier L, Topaz CM (2019) A topological approach to selecting models of biological experiments. *PLoS One* 14(3):e0213679
- Wasserman L (2018) Topological data analysis. *Annu Rev Stat Appl* 5:501–532
- Wolke U, Jezuit EA, Priess JR (2007) Actin-dependent cytoplasmic streaming in *C. elegans* oogenesis. *Development* 134(12):2227–2236

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Maria-Veronica Ciocanel¹  · Riley Juenemann² · Adriana T. Dawes³ · Scott A. McKinley²

¹ Department of Mathematics and Department of Biology, Duke University, Durham, USA

² Department of Mathematics, Tulane University, New Orleans, USA

³ Department of Mathematics and Department of Molecular Genetics, The Ohio State University, Columbus, USA