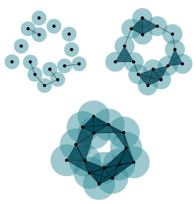
Research Statement

Rachel A. Neville

Two driving questions that arise in the study of complex spatiotemporal data are:

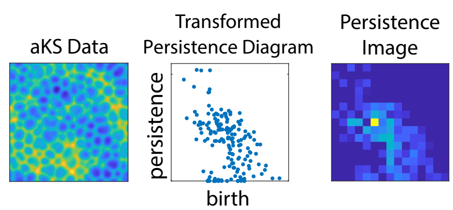
- How can one efficiently characterize and compare spatiotemporally complex patterns?
- How can changes in geometric and topological structure inform our understanding of the mechanisms driving the dynamics in the system?

I address these questions predominantly through persistent homology [5], a method from algebraic topology which detects and measures topological structure of data or functions by tracking changes in topological features (such as connected components, holes, or enclosed voids) through a filtration on the data. Three snapshots of a common distance-based filtration on point cloud data are shown in the figure below. The result of computing persistence is a multiscale representation of the topological structure of potentially very high dimensional data, and takes the form of a collection of sets of points.

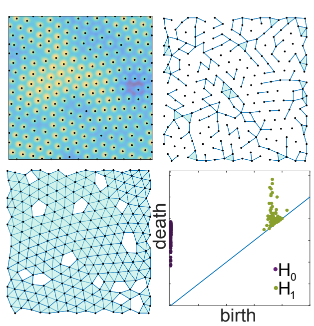


Flexibility in this framework requires careful design of filtrations to preserve interpretability and to capture discriminatory features of the data. Once persistent homology is computed, the results must be interpreted, that is, transformed to a space with additional structure (discussed in the next section) or summarized in a way that highlights features such as irregularities. My research focuses on developing filtrations as well as transformations and summaries of persistence diagrams in the context of studying partial differential equations and dynamical system models.

Connecting Persistence and Machine Learning We developed *persistence images* [3], a stable vector representation of persistence diagrams, shown in the figure on the right. This opens the door to a wide variety of previously unavailable machine learning techniques. This is instrumental in, for example, learning model parameters from samples (discussed later). Persistence images have been incorporated into several leading software packages for topological data analysis [4, 11].



In collaboration with a group from a Women in Computational Topology workshop, we refined a heat map representation of two simultaneous filtrations [2], which can give a richer representation of the data. Both of these tools are important for interpretation of persistence diagrams. In the following sections I will discuss a number of applications where transforming persistence diagrams or using them to build statistics on the data opened the door to a number of machine learning techniques and was vital to studying the system.



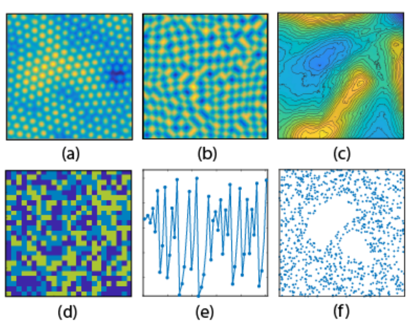
Quantifying Properties of Patterns Quantifying topological structure of a pattern leads to meaningful low dimensional summaries of data and can be employed to study geometric properties such as disorder, irregularities, or roughness of an evolving structure.

In [7], we defined a persistence-based measure of order for near hexagonal arrays of nanodots formed by ion bombardment of a binary compound. The bottom right panel in the figure on the left is the persistence diagram which quantifies defects in the pattern in the top left panel. Two examples in the filtration are shown, and highlight defects in the lattice pattern. The persistence-based measure captures disorder at a range of length scales, while other methods lacked sensitivity in certain contexts. This method can be extended to characterize general lattices, and defects therein, described as any integer combination of two vectors. Using this measure over the evolution of patterns, we examined the influence of a specific parameter in the long-term overall order. Well-ordered patterns take time to evolve. We found, surprisingly, that under certain conditions, a slower initial commitment to a pattern leads to better long-term order. We plan to use this method to study the conditions under which defects spontaneously resolve or catastrophically worsen.

For brevity, I will summarize several of my projects which characterize complex structure.

- With a team from the CSU Pattern Analysis Laboratory, we built a readily computable fractal dimension [1] for measures. Initial finding suggest that the distributions created from the persistence diagrams might be a useful *techniques for distinguishing between types of noise* (i.e. noise originating from different probability measures).
- With a group of experimental physicists in Helmholtz-Zentrum Dresden-Rossendorf in Germany, we developed a persistence based statistic that measures the prominence *structures that arise when a semiconductor is bombarded with an ion beam* at an elevated temperature [10], [8]. (See panel (b), below.)
- With a group in the CSU Watershed Science program, we developed a persistence based measure of the multiscale roughness of melting snowpack. This is used to *quantify roughness of snowfields* from LIDAR data [6] which affects wind resistance, energy exchange, and meltwater production. (See panel (c), below.)
- With colleagues at Oxford, Cambridge, and Berkeley (graduate student), we are using persistence to *measure "patchiness" of coral, algae turf, and micro-algae* and to compare coral reef data to the leading models. (See panel (d), below.)

Model and Parameter Learning Self-organized pattern formation is frequently the result of interactions of a number of different phenomenological processes which are encoded in the parameters of mathematical models. Some processes are well understood, and others are not. Insight into the influence of parameters helps build understanding of the underlying mechanisms driving the dynamics.



The next set of projects illustrate how persistence diagrams can be leveraged for parameter identification in simulated data. In [9], we use a signature in the persistence diagram to learn parameters of noisy data arising from the logistic map, a discrete time dynamical systems which exhibits chaotic behavior (see panel (e)). We leveraged persistence images and support vector machines to classify simulated data arising from the linked twist map, a discrete dynamical system that models mixing, by the parameter driving the system [3] (see panel (f)). Previously these types of classification were intractable. We also used persistence images and a subspace discriminant ensemble to learn parameters in data generated from the anisotropic Kuramoto-Sivashinsky (aKS) equation (a partial differential equation model), which exhibits a phenomena called chaotic bubbling. (See fig. on page 1.)

The Bradley-Shipman (BS) equations [12] which model ion bombardment of a binary compound have several linear and nonlinear parameters, which provides a richer context for investigating parameters. With persistence images, we performed similar classification tasks on various sets of parameter values and were able to distinguish between different linear and nonlinear parameters [8]. It is hypothesized that interactions between nonlinear parameters can cause defects to catastrophically worsen, but further investigation into the effects of these nonlinear parameters is needed. Along these lines, we are working to build a more robust regression model for learning parameters from data in this context and in a similarly posed problem, for the complex Ginzburg-Landau equations.

When we have access to real-world data, we use persistence to build a framework for comparing data with the associated model. This helps both to assess if mechanisms in the mathematical models accurately capture the phenomena displayed in the data as well as to learn parameter values from data. Initial results in the context of the two ion bombardment systems and the coral project (i.e. where data is available) the show promise. This is an area of ongoing inquiry.

Future Work Persistent homology is a flexible framework that is powerful for creating low-dimensional representations of essential features of complex data, characterizing geometric structure, learning underlying parameters and probing their influences, and comparing data to models. I am most interested in is developing methods for a more robust inclusion of temporal evolution of structures, rather than snapshots of their evolution. This will open the door to better understanding temporal changes in dynamic data.

My research offers many entry points to undergraduate and graduate students and attracts students with a wide range of interests, such as math, statistics, data science, physics, and biology. The applications of techniques developed in my research are wide-ranging. Recent software developments make the computation of persistence much more accessible. Relying on data to initially introduce a student to these mathematical techniques can motivate questions about both the data and the mathematics. I look forward to mentoring students on these projects.

References

[1] H. Adams, M. Aminian, E. Farnell, M. Kirby, C. Peterson, J. Mirth, R. Neville, and C. Shonkwiler. A fractal dimension for measures via persistent homology. *To appear in Proceedings of the Abel Symposia*, 2019.

[2] C. Betancourt, M. Chalifour, R. Neville, M. Pietrosanu, M. Tsuruga, I. Darcy, and G. Heo. *Pseudo-Multidimensional Persistence and Its Applications*, pages 179–202. Springer International Publishing, 2018.

[3] H. Adams S. Chepushtanova, T. Emerson, E. Hanson, M. Kirby, F. Motta, R. Neville, C. Peterson, P. Shipman, and L. Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18(8):1–35, 2017.

[4] Pawel Dlotko. Persistence representations. In *GUDHI User and Reference Manual*. GUDHI Editorial Board, 2017.

[5] H. Edelsbrunner and J. Harer. Persistent homology – a survey. *Contemporary Mathematics*, 453:257–282, 2008.

[6] S. Fassnacht, J. Kirkpatrick, R. Pasquini, R. Neville, P. Shipman, and G. Sextone. Geometric methods in the study of the snow surface roughness. *preprint*.

[7] F. C. Motta, R. Neville, P. D. Shipman, D. A. Pearson, and R. M. Bradley. Measures of order for nearly hexagonal lattices. *Physica D: Nonlinear Phenomena*, 380-381:17 – 30, 2018.

[8] R. Neville. *Topological Techniques for Characterization of Patterns in Differential Equations*. PhD thesis.

[9] R. Neville. Persistent homology of the logistic map: an exploration of chaos. Master's thesis, Colorado State University, 2014.

[10] X. Ou, A. Keller, M. Helm, J. Fassbender, and S. Facsko. Reverse epitaxy of Ge: Ordered and faceted surface patterns. *Physical Review Letters*, 111(1), 2013.

[11] Nathaniel Saul and Chris Tralie. Scikit-tda: Topological data analysis for python, 2019.

[12] P. D Shipman and R. M. Bradley. Theory of nanoscale pattern formation induced by normal-incidence ion bombardment of binary compounds. *Physical Review B*, 84(8):085420, 2011.