My research lies at the intersection of computer science, pure & applied mathematics, and statistics. Specifically, I work in the field of computational topology and geometry, and some of my work is in the sub-field of topological data analysis (TDA). I often colloquially describe my research as *finding shape and structure in data.* As a computational topologist, I often use algebraic structures to describe shape. We call these structures *topological descriptors.* In my research, I investigate questions such as: *How discerning are topological descriptors?* and *How do we compare shapes using distributions of descriptors?*

To accomplish my research goals, I take a collaborative and interdisciplinary approach. The develeopment of students and collaboration with peers is an integral aspect of my research. I have made novel contributions to my research community, have published in premier peer-reviewed venues, and have secured external funding to support my research activities. In the remainder of this document, I classify my current computational topology and geometry research into two themes: (1) The intersection of statistics and computational topology; and (2) Curves, graphs, and paths.

1 The Intersection of Statistics and Computational Topology

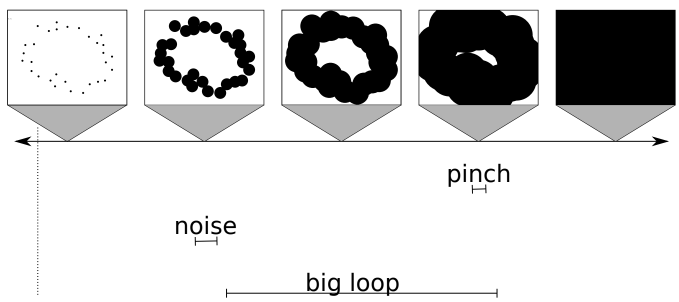


Figure 1: Given a finite point cloud, we construct a sequence of topological spaces by considering the union of balls of increasing radius, shown above from left to right. We think of the radius as increasing with time, and look at snapshots from the timeline above. The bars below the timeline are the *persistence barcode*, whose endpoints represented the *births* and *deaths* of topological features.

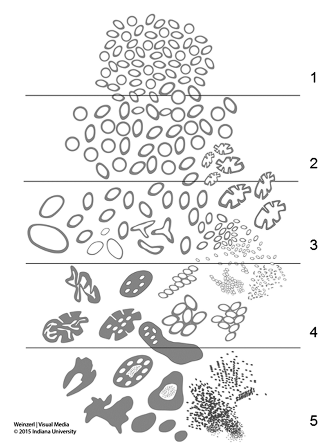
Before diving into my research, I want to provide a high-level introduction to a tool used in my research: persistent homology. Persistent homology is a method for studying the homology (i.e., the components, the tunnels, and the higher-dimensional 'voids') at multiple scales simultaneously. It provides a framework to quantify the evolution of the homology of a parameterized family of topological spaces. For example, we can study the persistent homology of a time-varying coverage region of a mobile sensor network, where one-dimensional features represent holes in the network coverage. Persistence tracks the homological changes that occur as the (time) parameter changes, pairing *births* or appearances of new features with *deaths* or merging of feature classes. This information is encoded in the *persistence diagram (or barcode)*, a multiset of points in the plane (a multiset of intervals in $\mathbb{R}$), each corresponding to the birth-death interval of some homological feature, as illustrated in Figure 1. Features that exist for a long interval can be viewed as topologically significant, while features with small intervals are indistinguishable from noise. The first theme of my research explores *how to use the persistence diagram and other topological descriptors in statistical and data analysis settings.*

Topological Descriptors as Statistics My annals of statistics paper [Fas+14] was among the first efforts combining statistics with computational topology. Since then, interest in statistical approaches to TDA has exploded, and using topological descriptors (e.g., persistence diagrams or Reeb graphs) is now more widey used as a data summary. The challenge with using persistence diagrams in statistical settings is that the space of persistence diagrams itself is ugly; for example, means are not unique. Thus, I investigate properties of spaces of topological descriptors [Buc+22; Fas+18a] and how to use them in statistical and data analysis settings [Ber+20; Cha+18; CK+22; Fas+20].

It is well-known that persistent homology is an invariant of the persistence module (that is, a sequence of vector spaces connected by linear operators formed from applying the homology functor to a filtered topological space), and not of the underlying filtered topological space itself. Thus, an inverse question that I investigate is: *how many filtrations over a fixed domain realize the same persistence diagram*? and *can we use a finite number of persistence diagrams to represent a given shape in Euclidean space*? In collaborations with others, I have results on the former question where the domain is a disk [Cat+20], as well as the latter question by developing algorithms to reconstruct the shape from directional topological descriptors [Bel+18; Bel+20; Fas+22a; Fas+18b]. This ties in to one of my current focuses in my research: how can we best understand a data set as a (potentially parameterized) distribution of topological descriptors (rather than just one descriptor)? I, along with collaborator Amit Patel, laid out some of the foundations for answering this question in a recent ArXiv preprint [FP22].

In addition to the persistence diagram, other topological descriptors are used throughout the literature, including: Betti curves, Euler characteristic curves, and persistence intensity plots. In [Ber+20], we included a survey of the common functional summaries of persistence diagrams. The benefit of functional summaries is that tools from functional data analysis can be leveraged. However, some information about the underlying filtration is lost. As such, one of my current investigations (that will become part of Anna Schenfisch's PhD dissertation) is to develop a framework for comparing topological descriptors.

R Package TDA One of my highly valued research products is the R package TDA.[1] Initially developed when I was a postdoc at CMU in collaboration with graduate student Fabrizio Lecci when writing the 2014 Annals paper, the R package now has 51,958 downloads to date (with 710 in October 2022). Instructors use this package in their classrooms,[2] and researchers use this package to analyze their data using topological techniques.[3]



Application: Digital Pathology The widespread availability of digital pathology images opens up new possibilities to use computational approaches to leverage the information inherent within them for diagnosis and prognosis. A collaboration with researchers at Tulane University aims to discover new quantitative image-based prognostic biomarkers (data descriptors) for prostate cancer. Since the structures used for prostate cancer grading are geometric and topological in nature, persistent homology is the natural tool to investigate; see image to the right, showing progression of the Gleason grades.[4] In this project, I have been active in the research surrounding searching in the space of persistence diagrams and topological descriptors, motivated by the need to classify biopsy images (or cropped sub-images). Searching in these spaces takes linear time (essentially, we need to compare a query diagram

[1] `http://cran.r-project.org/web/packages/TDA/index.html`

[2] See an assignment from a 2018-19 course on *Geometric Foundations of Data Analysis*: `http://www.maths.nuigalway.ie/~mstudies/2018-19/MA500/uploads/problems.pdf`.

[3] For example, see: Giansiracusa, Noah, Robert Giansiracusa, and Chul Moon. "Persistent homology machine learning for fingerprint classification," in *Proc. 18th IEEE Internat. Conf. Mach. Learn. (ICML)* 2019. Preprint available at arXiv:1711.09158

[4] Image from DOI: 10.1097/PAS.0000000000000530

against every diagram in a database). I have worked on an approximation algorithms based on provable guarantees [Fas+18a] by leveraging techniques from locality-sensitive hashing, using neural networks [Qin+21] by adapting image-searching machine learning methods, as well as developing a modified version of DBSCAN for topological descriptors [Fas+22b]. See `http://www.cs.tulane.edu/~carola/research/qubbd.html` for more details on this project, including a list of publications resulting from this work.

Selected Journal and Conference Publications

[Bel+18] R. L. Belton, B. T. Fasy, R. Mertz, S. Micka, D. L. Millman, D. Salinas, A. Schenfisch, J. Schupbach, and L. Williams. "Learning Simplicial Complexes from Persistence Diagrams". In: *Proc.* 30^{th} *Canadian Conf. Comput. Geom.* Also available at arXiv:1805.10716. U. Manitoba, Winnipeg, Canada, 2018.

[Bel+20] R. L. Belton, B. T. Fasy, R. Mertz, S. Micka, D. L. Millman, D. Salinas, A. Schenfisch, J. Schupbach, and L. Williams. "Reconstructing Embedded Graphs from Persistence Diagrams". In: *Computational Geometry: Theory and Applications* (2020). Also available at arXiv:1805.10716. DOI: `10.1016/j.comgeo.2020.101658`.

[Ber+20] E. Berry, Y.-C. Chen, J. Cisewski-Kehe, and B. T. Fasy. "Functional Summaries of Persistence Diagrams". In: *J. Appl. Comput. Topology (APCT)* 4 (2020). Also available at arXiv:1804.01618, pp. 211–262.

[Buc+22] M. Buchin, E. Chambers, B. T. Fasy, E. Gasparovic, E. Munch, and C. Wenk. "Comparing Distance Measures on Graphs". In: *La Matematica* (2022). To appear.

[Cat+20] M. J. Catanzaro, J. M. Curry, B. T. Fasy, J. Lazovskis, G. Malen, H. Riess, B. Wang, and M. Zabka. "Moduli Spaces of Morse Functions for Persistence". In: *J. Appl. Comput. Topology (APCT)* (2020). Preprint available at arXiv:1909.10623.

[Cha+18] F. Chazal, B. T. Fasy, F. Lecci, B. Michel, A. Rinaldo, and L. Wasserman. "Robust Topological Inference: Distance To a Measure and Kernel Distance". In: *J. Mach. Learn. Res.* 18 (159 2018), pp. 5845–84.

[CK+22] J. Cisewski-Kehe, B. T. Fasy, W. Hellwing, M. R. Lovell, P. Drozda, and M. Wu. "Differentiating Small-Scale Subhalo Distributions in CDM and WDM Models Using Persistent Homology". In: *Physical Review D* 106 (2022).

[Fas+14] B. T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, and A. Singh. "Confidence Sets For Persistence Diagrams". In: *Annals of Statistics* 42.6 (2014). Also available at ArXiv:1303.7117., pp. 2301–39.

[Fas+22a] B. T. Fasy, S. Micka, D. L. Millman, A. Schenfisch, and L. Williams. "Efficient Graph Reconstruction and Representation Using Augmented Persistence Diagrams". In: *Proc.* 34^{th} *Canadian Conf. Comput. Geom.* Also available at arXiv:1912.12759. 2022.

[Qin+21] Y. Qin, B. T. Fasy, C. Wenk, and B. Summa. "A Domain-Oblivious Approach for Learning Concise Representations of Filtered Topological Spaces". In: *IEEE Transactions on Visualization and Computer Graphics (VIS2021)* (2021). Also available at arxiv:2105.12208.

Selected Workshop Contributions, Posters, and Unpublished Work

[Fas+18a] B. T. Fasy, X. He, Z. Liu, S. Micka, M. D. L., and B. Zhu. "Approximate Nearest Neighbors in the Space of Persistence Diagrams". In submission to the AIMS journal *Foundations of Data Science*. Preprint available at arXiv:1812.1125. 2018.

[Fas+18b] B. T. Fasy, S. Micka, D. Millman, A. Schenfisch, and L. Williams. "Challenges in Reconstructing Shapes from Euler Characteristic Curves". In: *Proc. Fall Wksp. Comput. Geom.* Also available at arXiv:1811.11337. 2018.

[Fas+20] B. T. Fasy, S. Micka, D. Millman, and J. Wilson. *Too Many or Too Few? Sampling Bounds for Topological Descriptors*. In preparation. 2020.

[Fas+22b] B. T. Fasy, D. L. Millman, E. Pryor, and N. Stouffer. "DBSpan: Density-Based Spanner for Clustering Complex Data, With an Application to Persistence Diagrams". In: *Applications of Topological Data Analysis to Data Science, Artificial Intelligence, and Machine Learning (TDA at SDM)*. 2022.

[FP22] B. T. Fasy and A. Patel. "Persistent Homology Transform Cosheaf". ArXiv preprint arXiv:2208.05243. 2022.

2 Curves, Graphs, and Paths

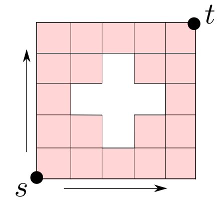
Extending some of the work that I started as a postdoc with Carola Wenk at Tulane [AFW14a; AFW14b; Ahm+15b], I work on several projects in the broad theme of reconstructing and comparing curves, graphs, and paths.

Comparing Embedded and Immersed Graphs My investigation of embedded and immersed graphs is motivated by the prevalence of GPS-tagged data that is closely tied to a road network (which can be modeled as a graph immersed in $\mathbb{R}^2$). I have worked directly on graph reconstruction [Ahm+15a; Buc+17; Fas+17; Fas+22], and on map-matching [Cha+20], where GPS trajectories are the input data and we wish to either reconstruct the underlying graph or to match the trajectory to a part of a known graph. In this research area, there is a need for a deeper study of embedded graphs and the distances between embedded graphs. When considering graph comparison from a graph-theoretical point of view, the obvious approaches to model graph comparison are NP-hard, such as subgraph isomorphism or graph edit distance. On the other hand, many computable distances require strong assumptions on the graphs. If the graphs represent road networks, these assumptions (such as planarity and minimum vertex degree) are unrealistic. With collaborators, I have proposed new road network comparisons for directed and weighted road networks [Bit+18], and developed methods to compare layered near-planar road networks (e.g., networks with bridges and tunnels) [Abd+19]. In fact, with REU student Emily Flanagan, we showed that computing an optimal layering is NP-hard [FFM20]. Recently, I led a group of computational topologists in writing a survey of metrics and distance measures for embedded and immersed graphs [Buc+22], and we are currently investigating the topological properties of these spaces.

Data on Graphs I am interested in studying data on graphs, and (more importantly) how to compare data on *different* graphs (e.g., census data by street address across years, where the graph structure of the streets changes). For example: inspired by conversations at Dagstuhl seminar 16022 (Geometric and Graph-based Approaches to Collective Motion), I proposed a research question to (now graduated) PhD student Sam Micka: *how do you efficiently monitor multiple flows on a single network*? We collaborated with three others at MSU to design an algorithm for placing monitors on a multi-flow network [Mic+16; Mic+17]. Two years after the mentioned Dagstuhl seminar, I co-organized the follow-up seminar 17282 entitled *From Observations to Prediction of Movement* [Bir+18].

Homotopy Area All closed curves in the plane are null homotopic. That is, they can continuously deform to a point. In collaboration with David Millman, Carola Wenk, and a few students, I have investigated algorithms for computing null homotopies that minimize area swept [FKW16; FKW17; Cha+22], as well as properties of a class of curves known as *self-overlapping (SO) curves* [EFW20]. SO curves are curves that are the boundary of an immersed disk, and we show how to decompose a curve into SO curves that realize the minimum area homotopy in [FKW17]. Bradley McCoy (one of my PhD students) is working on improving this algorithm and interpreting an algebraic construction of the existence of a null homotopy[5] as a geometric null homotopy, a nontrivial task.

[5] Zipei Nei. On the Minimum Area of Null Homotopies of Curves Traced Twice. `https://arxiv.org/abs/1412.0101`



Directed Cubical Complexes We model the state space for a concurrent program as follows: each thread of a concurrent program has its own time axis, and a cubical complex models all possible allowable states; see the *Swiss flag example* on the right. To study this directed complex, we study the spaces of directed paths with fixed endpoints, s and t. Such models come up in other settings as well, such as the parameter space for multi-parameter persistence. Then, a directed path from s to t is monotone with respect to all axes and represents an execution path of the program. My interest in directed topology started in November 2017 at an MSRI workshop for Women in Topology (WIT). I was assigned to the group led by Lisbeth Fajstrup focusing on "directed homotopy and homology theory, with an eye towards applications." After the workshop, I stepped up to co-lead the group of five graduate students and one other junior professor with Dr. Fajstrup. We published one paper where we defined a notion of directed collapsibility using past links and we proved a relationship between the directed past link of a directed cubical complex and connectedness and contractibility on the spaces of directed paths [Bel+20]. In our second paper, we investigated the relationships between directed collapsing and the preservation of directed path spaces [Bel+22]. Currently, we are working on leveraging our previous work in order to simplify parameter spaces for multiparameter persistence. In July 2023, Dr. Fajstrup and I will co-lead a WinCompTop research group on the connections between directed topology and dynamic programming at WinCompTop3.

Selected Journal and Conference Publications

[Ahm+15a] M. Ahmed, B. T. Fasy, M. Gibson, and C. Wenk. "Choosing Thresholds for Density-Based Map Construction Algorithms". In: *SIGSPATIAL*. ACM. 2015.

[AFW14a] M. Ahmed, B. T. Fasy, and C. Wenk. "Local Persistent Homology Based Distance Between Maps". In: *SIGSPATIAL*. ACM. 2014, pp. 43–52.

[Ahm+15b] M. Ahmed, B. T. Fasy, K. Hickmann, and C. Wenk. "A Path-Based Distance for Street Map Comparison". In: *Trans. Spatial Algorithms Syst.* 1.3 (1 2015). Also available at arXiv:1309.6131, 28 pages.

[Bel+20] R. Belton, R. Brooks, S. Ebli, L. Fajstrup, B. Fasy, C. Ray, N. Sanderson, and E. Vidaurre. "An Obstruction to Contractibility of Spaces of Directed Paths". In: *Advances in Mathematical Sciences*. Ed. by B. Acu, D. Danielli, M. Lewicka, A. Pati, S. Ramanathapuram Vancheeswara, and M. Teboh-Ewungkem. Association for Women in Mathematics Series. United States: Springer Publishing Company, 2020.

[Bel+22] R. Belton, R. Brooks, S. Ebli, L. Fajstrup, B. Fasy, N. Sanderson, and E. Vidaurre. "Combinatorial Conditions for Directed Collapsing". In: Association for Women in Mathematics Series. Springer Publishing Company, 2022.

[Bit+18] A. Bittner, B. T. Fasy, M. Grudzien, S. Ghosh, J. Huang, K. Pelatt, C. Thatcher, A. Tumurbaatar, and C. Wenk. "Comparing Directed and Weighted Road Maps". In: *Research in Computational Topology*. Ed. by E. Chambers, B. T. Fasy, and L. Ziegelmeier. AWM and IMA Series. Springer, 2018, pp. 57–70.

[Buc+17] K. Buchin, M. Buchin, D. Duran, B. T. Fasy, R. Jacobs, V. Sacristan, R. I. Silveira, F. Staals, and C. Wenk. "Clustering Trajectories for Map Construction". In: *ACM SIGSPATIAL GIS*. 2017.

[Buc+22] M. Buchin, E. Chambers, B. T. Fasy, E. Gasparovic, E. Munch, and C. Wenk. "Comparing Distance Measures on Graphs". In: *La Matematica* (2022). To appear.

[Cha+20] E. Chambers, B. T. Fasy, Y. Wang, and C. Wenk. "Map-Matching Using Shortest Paths". In: *Trans. Spatial Algorithms Syst.* 6.1 (2020), Article 6.

[EFW20] P. Evans, B. T. Fasy, and C. Wenk. "Combinatorial Properties of Self-Overlapping Curves and Interior Boundaries". In: *Proc. 36^{th} Annu. Symp. Comput. Geom.* 2020.

[Fas+22] B. T. Fasy, R. Komendarczýk, S. Majhi, and C. Wenk. "On the Reconstruction of Geodesic Subspaces of $\mathbb{R}^N$". In: *International Journal of Computational Geometry & Applications* 32.01n02 (2022), pp. 91–117.

[Mic+16] S. Micka, S. Yaw, B. T. Fasy, B. Mumey, and M. P. Wittie. "Efficient Multipath Flow Monitoring (presentation and abstract)". In: *International Highway Engineering Exchange Program (IHEEP)*. Second prize in student presentation. 2016.

[Mic+17] S. Micka, S. Yaw, B. T. Fasy, B. Mumey, and M. P. Wittie. "Efficient Multipath Flow Monitoring". In: *Proc. IFIP Networking*. 2017.

Selected Workshop Contributions, Posters, and Unpublished Work

[Abd+19] A. Abdelkader, G. Boeing, B. T. Fasy, and D. Millman. *Beyond Planarity: A Layered Representation for Transportation Networks*. In preparation. 2019.

[AFW14b] M. Ahmed, B. T. Fasy, and C. Wenk. "New Techniques in Road Network Comparison". In: *Grace Hopper Celebr. Women Comput.* Online proceedings. 2014.

[Bir+18] M. Birkin, S. Dodge, B. T. Fasy, and R. P. Mann. "From Observations to Prediction of Movement (Dagstuhl Seminar 17282)". In: *Dagstuhl Reports* 7.7 (2018), pp. 54–71. ISSN: 2192-5283. DOI: `10.4230/DagRep.7.7.54`. URL: `https://drops.dagstuhl.de/opus/volltexte/2018/8423`.

[Cha+22] H.-C. Chang, B. Fasy, B. McCoy, D. Millman, and C. Wenk. *From Curves to Words and Back Again: Geometric Computation of Minimum Area Homotopy*. In submission to SoCG 2023. 2022.

[FKW16] B. T. Fasy, S. Karakoc, and C. Wenk. "On Minimum Area Homotopies of Normal Curves in the Plane". In: Young Researcher's Forum of Computational Geometry Week. 2016.

[FKW17] B. T. Fasy, S. Karakoc, and C. Wenk. *On Minimum Area Homotopies of Normal Curves in the Plane*. Will be submitted to *Homology, Homotopy, and Applications*. ArXiv preprint arXiv:1707.02251. 2017.

[Fas+17] B. T. Fasy, R. Komendarczýk, S. Majhi, and C. Wenk. "Topological and Geometric Reconstruction of Metric Graphs in R^n". In: *Proc. Fall Wksp. Comput. Geom.* 2017, pp. 78–82.

[FFM20] E. Flanagan, B. T. Fasy, and D. Millman. *Optimal Partitioning of Road Networks*. NCUR abstract (not presented due to COVID-19). 2020.