My main area of research is topological data analysis (TDA). My work involves both applications and theory. On the application side, I use TDA to study the underlying structure of data sets. On the theory side, I study mathematical objects that are fundamental to TDA. I also work in complex social systems, where my research has focused on modeling opinion dynamics on networks.

Topological Data Analysis. The primary tool in TDA is persistent homology (PH), which is concerned with studying and quantifying the holes in data in all dimensions. This information is captured in a persistence diagram (PD), which details the scales at which holes form and fill in.

Applications: I have used PH to study geospatial data sets. In particular, I have used PH to study the accessibility of resources (such as polling sites, vaccine sites, and parks) in a geographic area to quantify areas with poor accessibility [9]. I plan to continue using TDA to study geospatial data sets, as well as expand my focus to other areas, especially in machine learning.

Theory: I have studied persistence modules, which are algebraic objects that are fundamental to PH, and have proved necessary and sufficient conditions for the existence of interval decompositions for a general class of persistence modules [11]. Interval decompositions of persistence module are the theoretical backbone in PH that provide the information in PDs. I plan to generalize my work on interval decompositions to wider classes of persistence modules.

Complex Social Systems. My work in this area has focused on studying the time evolution of opinions in social networks. I have formulated models of opinion dynamics and studied their behavior through mathematical analysis and numerical simulations [10]. I plan to continue these efforts by incorporating mean-field approaches and to model other social systems (such as disease spread on social networks).

1 Applications of Topological Data Analysis

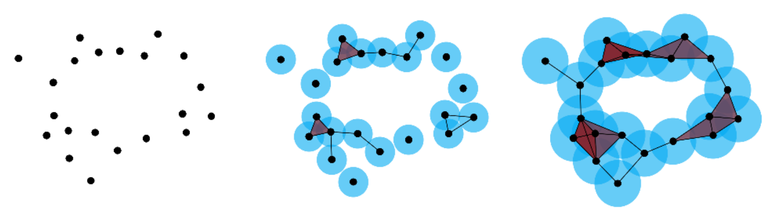

Figure 1: An example of a "Vietoris–Rips" filtration, which consists of simplicial complexes.

One of the primary tools in TDA is persistent homology (PH), which is concerned with the "persistence" of holes in data. Given a data set, often in the form of a point cloud, one constructs a *filtration*, which is a nested sequence $\mathcal{K}_1 \subseteq \mathcal{K}_2 \subseteq \cdots \subseteq \mathcal{K}_r$ of topological spaces that approximates the data set across different scales, and study how the homology of $\mathcal{K}_i$ changes as we increase i (the *filtration parameter*). By increasing i, new holes form and existing holes fill in. This information is summarized in a *persistence diagram* (PD), which is a multi-set $\{(x_i, y_i)\}_{i \in I}$ of points in $\mathbb{R}^2$; each point (x_i, y_i) corresponds to a hole, where x_i and y_i are the filtration-parameter values that the hole is formed (i.e., the birth filtration-parameter value) and is filled in (i.e., the death filtration-parameter value), respectively.

My work on applications of TDA has focused on geospatial data. My collaborators and I used PH to evaluate and compare equity in the geographic placement of resource centers (such as polling sites, vaccine sites, and parks) by determining and quantifying "holes in coverage," which are geographic areas that have poor resource coverage. One important benefit in using PH to evaluate resource accessibility is that it allows us to simultaneously consider holes in coverage at all scales, rather than determining them by choosing arbitrary cutoff thresholds. I plan to continue using PH to study geospatial data, as well as explore its use in other areas, such as machine learning and biology. I especially wish to explore incorporating PH into the framework of neural networks.

1.1 Past Work

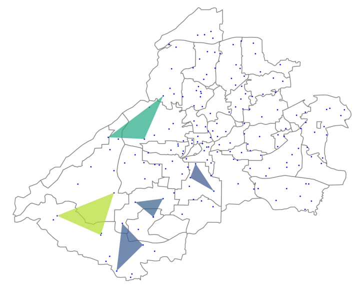

Figure 2: Death simplices representing holes in coverage in Atlanta.

1.1.1 Resource Coverage: Polling-Site Accessibility In [9], my collaborators[1] and I studied the accessibility of polling sites for five cities (Atlanta, Chicago, Jacksonville, New York City, and Salt Lake City) and Los Angeles County. For each of them, we constructed a weighted Vietoris–Rips (VR) filtration using estimates of average travel times, with weights based on waiting-time estimates at the polling sites.

1.2 Ongoing Work

1.2.1 Resource Coverage: Incorporating Resource Quality My collaborators[2] and I are building on the ideas in [9] to study the geographic distribution of resources with heterogeneous quality. In our applications, we incorporate not only distance to resources, but also the quality (represented by a scalar) of the resources. Different resource qualities (e.g., the available amenities of a park) lead to differences in how these resources affect nearby populations. To do this, we construct a bifiltration — a filtration that is indexed by two filtration parameters (which, in our case, are distance and quality) — and study the PH. The "resources" that we study are public parks, landfills, and pubs. Our choice in resources highlights the versatility in our methodology for measuring both desirable accessibility and unwanted exposure to a resource.

1.2.2 Resource Coverage: Assessing Solutions to Facility-Location Problems Another project (which I started at the American Mathematical Society's Mathematical Research Community on complex social systems) involves using PH to study facility-location problems (FLPs). FLPs [8] are optimization problems where, given a geographic area and its population distribution, one wishes to place facilities to serve a population by minimizing access cost. My collaborators[3] and I are using PH to evaluate and compare the equity in the placement of resources from solutions of FLPs by idenfifying and quantifying holes in coverage.

1.3 Future Work My work thus far in studying resource coverage with TDA has been limited to the distribution of resources at a fixed time. However, because distributions of resources can change with time, it is important to incorporate time into our TDA approaches. To use TDA to study resource accessibility in a time-dependent setting, I seek to explore the following approaches:

– <u>Zigzag Persistence</u> [4] is a generalization of PH that can be used for time-dependent data. A zigzag PD summarizes the times at which holes form and fill in.

– <u>Vineyards</u> [6] are time-dependent summaries of PH. Given time-dependent data, one can compute the PD for a fixed time. The associated vineyard is the continuously-varying stack of PDs.

Beyond resource-coverage problems, I hope to explore TDA applications in the following areas:

[1] My collaborators are Abigail Hickok (Columbia University), Benjamin Jarman (UCLA), Michael Johnson (UCLA), and Mason A. Porter (UCLA).

[2] My collaborators are Gillian Grindstaff (University of Oxford), Abigail Hickok (Columbia University), Sarah Tymochko (UCLA), and Mason A. Porter (UCLA).

[3] My collaborators are Giulia De Pasquale (ETH Zürich), Fabiana Ferracina (Washington State University), Rebecca Hardenbrook (Dartmouth College), Molly Lynch (Hollins University), J. Carlos Martínez Mori (MSRI), Anna C. Nelson (Duke University), Mason A. Porter (UCLA), and William Thompson (University of Delaware).

− TDA for Machine Learning: In summer 2022, I was an intern at the Pacific Northwest National Laboratory (PNNL), under the supervision of Gregory Henselman-Petrusek and Tegan Emerson, where I explored the pairing between vectorizations of PDs and machine-learning methods for classification tasks. I seek to further explore the use of TDA for machine learning, particularly in incorporating TDA into neural-network architectures (e.g., see [5]).

− Applications to Biology: I also plan to explore biological applications of TDA, such as in biological aggregation [13], cancer dynamics [12], and genomics [7].

1.4 Undergraduate Mentorship I will include undergraduate students in my research on applying TDA to real-world problems. The core ideas behind TDA and the ways that it is applied (which, at its core, predominantly requires understanding of linear algebra) are fairly accessible. Although I am also interested in working with students on other areas, resource-coverage problems (see Sections 1.1 and 1.2) are especially suitable for undergraduate involvement. Concrete ways that undergraduate students can contribute include studying new applications and adapting PH to capture relevant application-specific information. Undergraduate students who have experience (or wish to gain experience) in programming can also help with the coding aspects of projects. I am currently working with Amos Ancell on using PH to study the accessibility of fire stations. We will incorporate features (such as city boundaries) that are discussed in the "Future work" section (5.3) of [9] to construct our filtrations.

2 Theoretical Foundations of Topological Data Analysis

My work in theoretical TDA has focused on "persistence modules," which are algebraic objects that are fundamental to PH. Studying persistence modules builds on the theoretical backbone of TDA. Persistence modules also have important ramifications for practitioners, such as the relevance in the choice of field for computing PH. I plan to continue working on theoretical problems in TDA, especially those that are related to persistence modules.

Given a filtration $\mathcal{K} = \{\mathcal{K}_i\}_{i=1}^r$, one takes the kth homology with coefficients in a field F to obtain a *persistence module* $H_k(\mathcal{K}; F)$, which consists of (1) a sequence $\{H_k(\mathcal{K}_i; F)\}_{i=1}^r$ of vector spaces (which are called *homology groups*) over F that correspond to the topological spaces in a filtration and (2) linear maps (which are called *structure maps*) between homology groups that arise from inclusion. Using field coefficients to take homology guarantees the existence of an "interval decomposition" for the persistence module. Interval decompositions provides the necessary information to compute the corresponding PD.

Persistence diagrams depend on the choice of field. Choosing different fields to compute homology may yield different interval decompositions, which in turn yield different PDs. When computing homology with non-field coefficients (e.g., $\mathbb{Z}$-coefficients), interval decompositions may not exist.

2.1 Past Work One can study persistence modules of free Abelian groups to understand — from an algebraic perspective — when a filtration yields PDs that are independent of the choice of field. In [11], my collaborator Gregory Henselman-Petrusek (PNNL) and I studied conditions under which persistence modules of free Abelian groups admit an interval decomposition. Using tools from lattice theory, we built our framework to prove the following necessary and sufficient condition for the existence of interval decompositions.

Theorem 1. *A persistent module of free Abelian groups has an interval decomposition if and only if the cokernel of every induced map is free.*

We also developed a polynomial-time algorithm that produces an interval decomposition for persistence modules of free Abelian groups, provided one exists.

2.2 Ongoing and Future Work Gregory Henselman-Petrusek and I are currently building on our work on persistence modules of free Abelian groups. While the persistence modules in [11] are finitely indexed, we are generalizing our results to the setting of continuously-indexed (e.g., by $\mathbb{R}$ or $[0, 1]$) persistence modules. I expect that many of our ideas can be adapted to continuous settings. We are also extending our results to zigzag persistence modules. In particular, we are building a framework to study zigzag persistence modules of free Abelian groups and determining the conditions under which they admit interval decompositions.

We also plan to extend our work beyond the setting of free Abelian groups to general Abelian groups (i.e., including torsion). I expect that the main obstruction to overcome for this work will be the torsion of the groups in the persistence modules. This will likely require case-wise treatment for different orders of torsion.

2.3 Undergraduate Mentorship Theoretical projects in TDA offer students a chance to dive into theory that is relevant to and motivated by applications. Although they require more background than projects in my other research areas, they are accessible to motivated students who are proficient in algebra and topology. They also offer students a taste in theoretical fields (mainly algebra and topology) while requiring less background than other topics in these fields. They are especially well suited to students who are unsure whether they want to pursue pure or applied math, as well as students who wish to explore the intertwining between theory and application.

3 Complex Social Systems

In addition to studying resource accessibility (see Sections 1.1 and 1.2), my work in complex social systems includes modeling the time evolution of opinions in social networks. Formulating and studying models of opinion dynamics help us understand the factors that influence opinion changes. I have formulated and studied bounded-confidence models (BCMs), which is a family of opinion models in which agents (represented by nodes of a network) can only influence neighbors whose opinions are "close enough" (more precisely, when their opinions differ by less than a "confidence bound"). I plan to continue analyzing models of both opinion dynamics and other social systems.

3.1 Past Work

3.1.1 BCMs of Opinion Dynamics with Adaptive Confidence Bounds People tend to be more receptive to those with whom they interact positively. To model this, my collaborators[4] and I formulated and analyzed discrete-time BCMs with heterogeneous and adaptive confidence bounds [10], based on the idea that positive (respectively, negative) interactions between individuals increase (respectively, decrease) mutual receptiveness. We analytically and numerically explored our adaptive BCMs' long-term behaviors, including the confidence-bound dynamics, the formation of clusters of nodes with similar opinions (which we call "limiting opinion clusters"), and the time evolution of an "effective graph," which is a time-dependent subgraph with edges between nodes that can currently influence each other. We proved the following theorem, which quantifies the relationship between the effective graph and formation of limiting clusters.

[4]My collaborators are Grace J. Li (UCLA) and Mason A. Porter (UCLA).

Theorem 2. *The effective graph $G_{\text{eff}}(t)$ is eventually constant in time. Moreover, the edges in the* limiting effective graph $\lim_{t\to\infty} G_{\text{eff}}(t)$ *are between nodes in the same limiting opinion cluster.*

For a wide range of parameters that control the increase and decrease of confidence bounds, we demonstrated for a variety of networks that our adaptive BCMs result in fewer large opinion clusters and longer convergence times than associated baseline (i.e., nonadaptive) BCMs. We also demonstrate numerically that our models have structurally richer effective graphs than the baseline models.

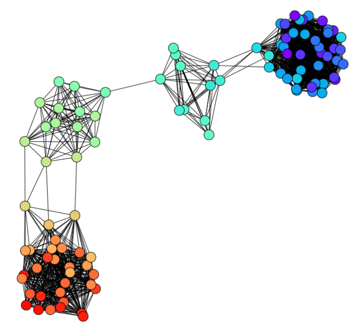

Figure 3: An effective graph, which has edges between mutually receptive nodes.

3.2 Ongoing Work

3.2.1 Coupled Disease and Opinion Dynamics To study the effects that opinions and disease spread have on each other, my collaborators[5] and I have proposed a multilayer network model that combines a discrete-time BCM with a discrete-time compartmental disease model. In this model, each agent has both an opinion and a disease state (susceptible, infected, or recovered), which influence each other. We will use mathematical analysis and numerical simulations to study the time evolution of the distribution of opinions, the magnitude and speed of disease spread, and the relationship between the two.

3.2.2 BCMs of Coupled Opinion Dynamics Individuals have opinions on different topics, and the dynamics of these opinions are often interdependent. To model this, my collaborators[6] and I have formulated and are currently analyzing a multi-opinion BCM in which agents have opinions on a variety of different topics. In our multi-opinion BCM, whether or not two agents compromise on one topic depends not only on their opinion difference for that topic, but also on their opinions on all other topics. We are numerically studying how the long-term behavior in our model differs from a baseline model with independently evolving opinions.

3.3 Future Work Agent-based models, including the models of opinion dynamics that I have studied, require Monte Carlo simulations — which often are computationally intensive and time consuming — to obtain robust statistics of long-term behavior. One can instead model opinion dynamics from a mean-field perspective by considering a density description of opinion distribution that evolve following an integro-differential equation (e.g., see [2]), which reduces the need to perform Monte Carlo simulations. I plan to use mean-field approaches to complement my work in agent-based modeling. Doing so will help with computation time and provide analytical insights into the behavior of associated agent-based models.

I also hope to model other social systems, such as disease dynamics on networks [3] and human mobility [1]. I hope to understand the ways in which different social systems can influence each other by coupling different social models together (in a similar spirit to Section 3.2.2).

[5] My collaborators are Yang Yang (The Ohio State University), Mason A. Porter (UCLA), and Joseph Tien (The Ohio State University).

[6] My collaborators are Grace J. Li (UCLA), Weiqi Chu (University of Massachusetts, Amherst), and Mason A. Porter (UCLA).

3.4 Undergraduate Mentorship I will involve undergraduate students in my research in complex social systems. Many projects in complex social systems (for example, projects on modeling opinion dynamics) are both down to earth and suitable for undergraduate involvement (often only requiring predominantly a basic understanding of linear algebra and numerical methods to get started), while also producing meaningful research that is of interest to expert practitioners. Many projects focus entirely on numerics, in which the required background is minimal and students can benefit by gaining valuable programming experience. Undergraduate students can contribute to these projects in many ways, including proposing new models, designing and running numerical simulations, and even proving theoretical results. In the past, I have formulated and studied a BCM that incorporates opinion repulsion with Xiaohe Zhang. I am currently investigating the behavior of a BCM on directed configuration-model networks with Ruyi Lu.

References

[1] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.

[2] E. Ben-Naim, P. L. Krapivsky, and S. Redner. Bifurcations and patterns in compromise processes. *Physica D: Nonlinear Phenomena*, 183(3):190–204, 2003.

[3] F. Brauer, C. Castillo-Chavez, and Z. Feng. Simple compartmental models for disease transmission. In *Mathematical Models in Epidemiology*, Texts in Applied Mathematics, pages 21–61. Springer, New York, NY, 2019.

[4] G. Carlsson and V. de Silva. Zigzag persistence. *Foundations of Computational Mathematics*, 10(4):367–405, 2010.

[5] M. Carriere, F. Chazal, Y. Ike, T. Lacombe, M. Royer, and Y. Umeda. PersLay: A neural network layer for persistence diagrams and new graph topological signatures. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pages 2786–2796. Proceedings of Machine Learning Research, 2020.

[6] D. Cohen-Steiner, H. Edelsbrunner, and D. Morozov. Vines and vineyards by updating persistence in linear time. In *Proceedings of the Twenty-Second Annual Symposium on Computational Geometry*, pages 119–126, Sedona, Arizona, USA, 2006. Association for Computing Machinery.

[7] P. G. Cámara. Topological methods for genomics: Present and future directions. *Current Opinion in Systems Biology*, 1:95–101, 2017.

[8] C. Filippi, G. Guastaroba, D. Huerta-Muñoz, and M. Speranza. A kernel search heuristic for a fair facility location problem. *Computers & Operations Research*, 132:105292, 2021.

[9] A. Hickok, B. Jarman, M. Johnson, J. Luo, and M. A. Porter. Persistent homology for resource coverage: A case study of access to polling sites, 2023. arXiv:2206.04834.

[10] G. J. Li, J. Luo, and M. A. Porter. Bounded-confidence models of opinion dynamics with adaptive confidence bounds, 2023. arXiv:2303.07563.

[11] J. Luo and G. Henselman-Petrusek. Interval decompositions for persistence modules of free abelian groups, 2023. arXiv:2310.07971.

[12] H. Masoomy, B. Askari, S. Tajik, A. K. Rizi, and G. R. Jafari. Topological analysis of interaction patterns in cancer-specific gene regulatory network: Persistent homology approach. *Scientific Reports*, 11:16414, 2021.

[13] C. M. Topaz, L. Ziegelmeier, and T. Halverson. Topological data analysis of biological aggregation models. *PLOS ONE*, 10(5):e0126383, 2015.